(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 October 2004 (21.10.2004)

PCT

(10) International Publication Number
WO 2004/090551 A2

(51) International Patent Classification[7]: G01N 33/68

(21) International Application Number:
PCT/EP2004/003737

(22) International Filing Date: 7 April 2004 (07.04.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/461,558	8 April 2003 (08.04.2003)	US
60/461,623	8 April 2003 (08.04.2003)	US
60/471,479	16 May 2003 (16.05.2003)	US
60/474,863	30 May 2003 (30.05.2003)	US
60/484,140	30 June 2003 (30.06.2003)	US

(71) Applicant *(for all designated States except US)*: **GENOVA, LTD.**; Canon's Court, 22 Victoria Street, Hamilton, HM12 (BM).

[AT/AT]; Brunner Strasse 59, I-1230 Vienna (AT).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ARGOUD-PUY, Guilaine** [FR/CH]; c/o GeneProt, Inc., 2, Pre-de-la-Fontaine, 1217 Meyrin (CH). **BEDERR, Nassima** [FR/CH]; c/o GeneProt, Inc., 2, Pre-de-la-Fontaine, CH-1217 Meyrin (CH). **BOUGUELERET, Lydie** [FR/CH]; c/o GeneProt, Inc., 2, Pre-de-la-Fontaine, CH-1217 Meyrin (CH). **CUSIN, Isabelle** [FR/CH]; c/o GeneProt, Inc., 2, rue Pre-de-la-Fontaine, CH-1217 Meyrin (CH). **MAHE, Eve** [FR/CH]; c/o GeneProt, Inc., 2, Pre-de-la-Fontaine, CH-1217 Meyrin (CH). **NIKNEJAD, Anne** [CH/CH]; c/o GeneProt, Inc., 2, Pre-de-la-Fontaine, CH-1217 Meyrin (CH). **REFFAS, Samia** [CH/CH]; c/o GeneProt, Inc., 2, Pre-de-la-Fontaine, CH-1217 Meyrin (CH).

(74) Agent: **GRUBB, Philip**; Novartis AG, Corporate Intellectuel Property, CH-4002 Basel (CH).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SECRETED POLYPEPTIDE SPECIES ASSOCIATED WITH CARDIOVASCULAR DISORDERS

(57) Abstract: The invention discloses human secreted polypeptides that circulate at an increased level in the plasma of patients with cardiovascular disorders. The invention also provides methods of using compositions including the polypeptides, polynucleotides encoding them, and antibodies specific for these polypeptides, for diagnosis, prognosis, and for drug development.

WO 2004/090551 A2

SECRETED POLYPEPTIDE SPECIES ASSOCIATED WITH CARDIOVASCULAR DISORDERS

FIELD OF THE INVENTION

The invention relates to polypeptide species secreted preferentially in individuals with cardiovascular disorders, to isolated polynucleotides encoding such polypeptides, to polymorphic variants thereof, and to the use of said nucleic acids and polypeptides or compositions thereof in detection assays, for cardiovascular disorder diagnosis, and for drug development.

BACKGROUND

Cardiovascular disease is a major health risk throughout the industrialized world. Coronary Artery Disease (CAD) is characterized by atherosclerosis or hardening of the arteries. Atherosclerosis is the most prevalent of cardiovascular diseases, is the principal cause of heart attack, stroke, and gangrene of the extremities, and thereby the principle cause of death in the United States. Atherosclerosis is a complex disease involving many cell types and molecular factors (described in, for example, Ross, 1993, Nature 362: 801-809). In normal circumstances a protective response to insults to the endothelium and smooth muscle cells (SMCs) of the wall of the artery consists of the

The advanced lesions of atherosclerosis may occlude the artery concerned, and result from an excessive inflammatory-fibroproliferative response to numerous different forms of insult. Injury or dysfunction of the vascular endothelium is a common feature of many conditions that predispose an individual to accelerated development of atherosclerotic cardiovascular disease.

Atherosclerotic plaques occlude the blood vessel concerned and restrict the flow of blood, resulting in ischemia. Ischemia is a condition characterized by a lack of oxygen supply in tissues of organs due to inadequate perfusion. Such inadequate perfusion can have a number of natural causes, including atherosclerotic or restenotic lesions, anemia, or stroke. The most common cause of ischemia in the heart is atherosclerotic disease of epicardial coronary arteries. By reducing the lumen of these vessels, atherosclerosis causes an absolute decrease in myocardial perfusion in the basal state or limits appropriate increases in perfusion when the demand for flow is augmented. Coronary blood flow can also be limited by arterial thrombi, spasm, and, rarely, coronary emboli, as well as by ostial narrowing due to luetic aortitis. Congenital abnormalities, such as anomalous origin of the left anterior descending coronary artery from the pulmonary artery, may cause myocardial ischemia and infarction in infancy, but this cause is very rare in adults.

Myocardial ischemia can also occur if myocardial oxygen demands are abnormally increased, as in severe ventricular hypertrophy due to hypertension or aortic stenosis. The latter can be present with angina that is indistinguishable from that caused by coronary atherosclerosis. A reduction in the oxygen-carrying capacity of the blood, as in extremely severe anemia or in the presence of carboxy-hemoglobin, is a rare cause of myocardial ischemia. Not infrequently, two or more causes of ischemia will coexist, such as an increase in oxygen demand due to left ventricular hypertrophy and a reduction in oxygen supply secondary to coronary atherosclerosis.

Extensive clinical studies have identified factors that increase the risk of cardiovascular disorders. Some of these risk factors, such as age, gender, and family history cannot be changed. Other risk factors include the following: smoking, high blood pressure, high fat and high cholesterol diet, diabetes, lack of exercise, obesity, and stress.

Fortunately, many contributing factors are controllable through lifestyle changes. The risk of cardiovascular disorders for smokers is more than twice that of non-smokers. When a person stops smoking, regardless of how much he or she may have smoked in the past, their risk of developing a disorder rapidly declines. Serum cholesterol level is directly related to prevalence of cardiovascular disorder and hypertension or high blood pressure is an important risk factor. Physical activity has been postulated to reduce the risk of developing a cardiovascular disorder through various

myocardial contraction and its electrical impulse stability. Reduced oxygen demand and myocardial work are reflected in lowered heart rate and blood pressure at rest. Physical activity also increases the diameter and dilatory capacity of coronary arteries, increases collateral artery formation, and reduces rates of progression of coronary artery atherosclerosis. Obesity and the serum fatty acids are reduced by activity.

There may be no noticeable symptoms of a cardiovascular disorder at rest, but symptoms such as chest pressure may occur with increased activity or stress. Other first signs that can appear are heartburn, nausea, vomiting, numbness, shortness of breath, heavy cold sweating, unexplained fatigue, and feelings of anxiety. The more severe symptoms of cardiovascular disorders are chest pain (angina pectoris), rhythm disturbances (arrhythmias), stroke, or heart attack (myocardial infarction). Strokes and heart attacks result from a blocked artery in the brain and heart tissue, respectively. Because symptoms vary, the tests and treatments chosen can be very different from one patient to another.

Diagnostic tests useful in determining the extent and severity of cardiovascular disorder include: electrocardiogram (EKG), stress test, nuclear scanning, coronary angiography, resting EKG, EKG Multiphase Information Diagnosis Indexes, Holter monitor, late potentials, EKG mapping,

echocardiogram, Thallium scan, PET, MRI, CT, angiogram and IVUS. Additional risk factor measures and useful diagnostics are common and best applied by one of skill in the art of medicine. There are many different therapeutic approaches, depending on the seriousness of the disease. For many people, cardiovascular disorders are managed with lifestyle changes and medications. More severe diagnoses may indicate a need for surgery.

Surgical approaches to the treatment of ischemic atherosclerosis include bypass grafting, coronary angioplasty, laser angioplasty, atherectomy, endarterectomy, and percutaneous translumenal angioplasty (PCTA). The failure rate after these approaches due to restenosis, in which the occlusions recur and often become even worse, is extraordinarily high (30-50%). It appears that much of the restenosis due to further inflammation, smooth muscle accumulation, and thrombosis. Additional therapeutic approaches to cardiovascular disease have included treatments that encouraged angiogenesis in such conditions as ischemic heart and limb disease.

The non-specific nature of most CAD and cardiovascular disorder symptoms makes definitive diagnosis difficult. More quantitative diagnostic methods suffer from variability, both between individuals and between readings on a single individual. Thus, diagnostic measures must be standardized and applied to individuals with well-documented and extensive medical histories. Further, current diagnostic methods often do not reveal the underlying cause for a given observation address the causative problem and may even be harmful to the individual.

Methods of diagnosis that rely on nucleotide detection include genetic approaches and expression profiling. For example, genes that are known to be involved in cardiovascular disorders may be screened for mutations using common genotyping techniques such as sequencing, hybridization-based techniques, or PCR. In another example, expression from a known gene may be tracked by standard techniques including RTPCR, various hybridization-based techniques, and sequencing. These strategies often do not enable a practitioner to detect differences in mRNA processing and splicing, translation rate, mRNA stability, and posttranslational modifications such as proteolytic processing, phosphorylation, glycosylation, and amidation.

To address the current weaknesses in the diagnostic state of the art for cardiovascular disorders, the invention provides specific plasma polypeptides that are differentially increased in plasma from individuals with Coronary Artery Disease compared to control plasma. By providing the actual polypeptide species, differences in mRNA processing and splicing, translation rate, mRNA stability, and posttranslational modifications such as proteolytic processing, phosphorylation, glycosylation, and amidation are revealed. The polypeptides of the invention are thus described as

"Cardiovascular disorder Plasma Polypeptides" or CPPs. These polypeptide sequences are described as SEQ ID NOs:1-2, 6-7, 11-12, 15-17, and 24-25, and those comprising at least one of the amino acid sequences selected from SEQ ID NOs:3-5, 8-10, 13-14, 18-23 and 26-28 (see Figures 1-5). These polypeptides are more precisely referred to as "Cardiovascular disorder Plasma Polypeptides 2, 9, 17, 20 and 21" (CPP 2 corresponding to SEQ ID NOs:1-5, CPP 9 corresponding to SEQ ID NOs:6-10, CPP 17 corresponding to SEQ ID NOs:15-23, CPP 20 corresponding to SEQ ID NOs:24-28 and CPP 21 corresponding to SEQ ID NOs:11-14), and include fragments, and post-translationally modified species of CPPs that are present at a higher level in plasma obtained from individuals with Coronary Artery Disease (CAD). Preferred fragments of the invention are those described as SEQ ID NOs:3-5, 8-10, 13-14, 18-23 and 26-28. Thus, the CPPs of the invention represent an important diagnostic tool for determining the risk of coronary artery disease (CAD), coronary heart disease (CHD), peripheral vascular disease, cerebral ischemia (stroke), congestive heart failure, atherosclerosis, hypertension, and other cardiovascular diseases. CPPs are secreted factors and as such, are readily detectable and useful for drug development, diagnosis, and prevention of cardiovascular diseases.

SUMMARY OF THE INVENTION

The present invention is directed to compositions related to secreted polypeptide species that

polypeptide species are designated herein "Cardiovascular disorder Plasma Polypeptides," or CPPs. Such Cardiovascular disorder Plasma Polypeptides comprise an amino acid sequence selected from one of the groups consisting of SEQ ID NOs:1-5, 6-10, 11-14, 15-23 and 24-28, and are designated CPP 2, CPP 9, CPP 17, CPP 20 and CPP 21, respectively. Compositions include CPP precursors, antibodies specific for CPPs, including monoclonal antibodies and other binding compositions derived therefrom. Further included are methods of making and using these compositions. Precursors of the invention include unmodified precursors, proteolytic precursors of SEQ ID NOs:1-28, and intermediates resulting from alternative proteolytic sites in the amino acid sequences of SEQ ID NOs:1-28.

A preferred embodiment of the invention includes CPPs having a posttranslational modification, such as a phosphorylation, glycosylation, acetylation, amidation, or a C-, N- or O-linked carbohydrate group. Additionally preferred are CPPs with intra- or inter-molecular interactions, e.g., disulfide and hydrogen bonds that result in higher order structures. Also preferred are CPPs that result from differential mRNA processing or splicing. Preferably, the CPPs represent post-translationally modified species, structural variants, or splice variants that are present in plasma

from individuals with a cardiovascular disorder.

In another aspect, the invention includes CPPs comprising a sequence which is at least 75 percent identical to a sequence selected from one of the groups consisting of SEQ ID NOs:1-5, and 11-23. Preferably, the invention includes polypeptides comprising at least 85 percent, and more preferably at least 90 percent, and still more preferably at least 95 percent, identity with any one of the sequences selected from SEQ ID NOs:1-5, and 11-23. Most preferably, the invention includes polypeptides comprising a sequence at least 99 percent identical to a sequence selected from one of the groups consisting of SEQ ID NOs:1-5, and 11-23.

In another aspect, the invention includes CPPs comprising a sequence which is at least 85 percent identical to a sequence selected from the group consisting of SEQ ID NOs:6-10. Preferably, the invention includes polypeptides comprising at least 90 percent, and more preferably at least 95 percent, and still more preferably at least 97 percent, identity with any one of the sequences selected from SEQ ID NOs:6-10. Most preferably, the invention includes polypeptides comprising a sequence at least 99 percent identical to a sequence selected from the group consisting of SEQ ID NOs:6-10.

In another aspect, the invention includes CPPs comprising a sequence which is at least 95 percent identical to a sequence selected from the group consisting of SEQ ID NOs:24-28. Preferably, the invention includes polypeptides comprising at least 97 percent, and more preferably at least 98 from SEQ ID NOs:24-28. Most preferably, the invention includes polypeptides comprising a sequence at least 99 percent identical to a sequence selected from the group consisting of SEQ ID NOs:24-28.

In another aspect, the invention includes natural variants of CPPs having a frequency in a selected population of at least two percent. More preferably, such natural variant has a frequency in a selected population of at least five percent, and still more preferably, at least ten percent. Most preferably, such natural variant has a frequency in a selected population of at least twenty percent. The selected population may be any recognized population of study in the field of population genetics. Preferably, the selected population is Caucasian, Negroid, or Asian. More preferably, the selected population is French, German, English, Spanish, Swiss, Japanese, Chinese, Irish, Korean, Singaporean, Icelandic, North American, Israeli, Arab, Turkish, Greek, Italian, Polish, Pacific Islander, Finnish, Norwegian, Swedish, Estonian, Austrian, or Indian. More preferably, the selected population is Icelandic, Saami, Finnish, French of Caucasian ancestry, Swiss, Singaporean of Chinese ancestry, Korean, Japanese, Quebecian, North American Pima Indians, Pennsylvanian Amish and Amish Mennonite, Newfoundlander, or Polynesian.

A preferred aspect of the invention provides a composition comprising an isolated CPP, i.e., a CPP free from proteins or protein isoforms having a significantly different isoelectric point or a significantly different apparent molecular weight from the CPP. The isoelectric point and molecular weight of a CPP may be indicated by affinity and size-based separation chromatography, 2-dimensional gel analysis, and mass spectrometry.

In a preferred aspect, the invention provides particular polypeptide species that comprise an amino acid sequence selected from the one of the groups consisting of SEQ ID NOs:3-5, 8-10, 13-14, 18-23 and 26-28. Preferably, the particular polypeptide species further comprises contiguous amino acid sequence from SEQ ID NOs:1-2, 6-7, 11-12, 15-17, and 24-25, respectively. Preferred species are polypeptides that i) comprise an amino acid sequence selected from one of the groups consisting of SEQ ID NOs: 3-5, 8-10, 13-14, 18-23 and 26-28; ii) appear at a higher level in plasma from individuals with a cardiovascular disorder; and iii) optionally result from proteolytic processing of the polypeptides of SEQ ID NO: 1-2, 6-7, 11-12, 15-17, and 24-25, respectively.

In a still further aspect, the invention provides a combination of two or more of the polypeptides selected from the one of the groups consisting consisting of SEQ ID NOs:1-5, 6-10, 11-14, 15-23 and 24-28.

In an additional aspect, the invention includes modified CPPs. Such modifications include labels, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the protein. Chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH4, acetylation, formylation, oxidation, reduction, or metabolic synthesis in the presence of tunicamycin.

Also provided by the invention are chemically modified derivatives of the polypeptides of the invention which may provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (e.g., water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol). The CPPs are modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

In another embodiment, the invention provides a method of identifying a modulator of at least one CPP biological activity comprising the steps of: i) contacting a test modulator of a CPP biological activity with the polypeptide comprising the amino acid sequence selected from one of the groups

consisting of the amino acid sequences listed in Table 1 (corresponding to CPP 2, CPP 9, CPP 17, CPP 20 and CPP 21); ii) detecting the level of said CPP biological activity; and iii) comparing the level of said CPP biological activity to that of a control sample lacking said test modulator. Where the difference in the level of CPP protein biological activity is a decrease, the test modulator is an inhibitor of at least one CPP biological activity. Where the difference in the level of CPP biological activity is an increase, the test substance is an activator of at least one CPP biological activity. For CPP 2, the CPP biological activity tested is preferably increasing / decreasing lipase activity.

Another aspect of the invention relates to a method of identifying a modulator of a cardiovascular disorder comprising the steps of: (a) administering a candidate agent to a non-human test animal which is predisposed to be affected or which is affected by the cardiovascular disorder; (b) administering the candidate agent of (a) to a matched control non-human animal not predisposed to be affected or not being affected by the cardiovascular disorder; (c) detecting and /or quantifying the level of at least one polypeptide in a biological sample obtained from the non-human test of (a) or control animal of (b), wherein the at least one polypeptide is selected from: (i) a polypeptide comprising the amino acid sequence selected from one of the groups consisting of SEQ ID NOs:1-2, 6-7, 11-12, 15-17, and 24-25; (ii) a variant, with at least 75% sequence identity, having one or more amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID

amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID NOs:6, or 7; (iv) a variant, with at least 95% sequence identity, having one or more amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID NOs:24, or 25; and (v) a fragment of a polypeptide as defined in i), ii), iii), or iv) above which is a least ten amino acids long; and step (d) comparing the level of the at least one polypeptide of step (c); wherein an alteration in the level of the polypeptide indicates that the candidate agent is a modulator of the cardiovascular disorder. A preferred embodiment of the invention provides that the non-human test animal which is predisposed to be affected or which is affected by the cardiovascular disorder comprises an increased plasma level of at least one of the polypeptides selected from: (i) a polypeptide comprising the amino acid sequence selected from one of the groups consisting of SEQ ID NOs:1-2, 6-7, 11-12, 15-17, and 24-25; (ii) a variant, with at least 75% sequence identity, having one or more amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID NOs:1, 2, 11, 12, 15, 16, or 17; (iii) a variant, with at least 85% sequence identity, having one or more amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID NOs:6, or 7; (iv) a variant, with at least 95% sequence identity, having

one or more amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID NOs:24, or 25; and (v) a fragment of a polypeptide as defined in i), ii), iii), or iv) above which is a least ten amino acids long.

In a still further aspect of the invention, a method for monitoring the efficacy of a treatment of a subject having or at risk of developing a cardiovascular disorder with an agent is provided, which comprises steps: (a) obtaining a pre-administration biological sample from the subject prior to administration of the agent; (b) detecting and /or quantifying the level of at least one polypeptide in the biological sample from said subject, wherein the at least polypeptide is selected from: (i) a polypeptide comprising the amino acid sequence selected from one of the groups consisting of SEQ ID NOs:1-2, 6-7, 11-12, 15-17, and 24-25; (ii) a variant, with at least 75% sequence identity, having one or more amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID NOs:1, 2, 11, 12, 15, 16, or 17; (iii) a variant, with at least 85% sequence identity, having one or more amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID NOs:6, or 7; (iv) a variant, with at least 95% sequence identity, having one or more amino acid substitutions, deletions or insertions relative to the amino acid sequence shown in SEQ ID NOs:24, or 25; and (v) a fragment of a polypeptide as defined in i), ii), iii), or iv) above which is a least ten amino acids long; and steps (c) obtaining one or more post-administration administration sample or samples; (e) comparing the level of the at least one polypeptide in the pre-administration sample with the level of the at least one polypeptide in the post- administration sample; and (f) adjusting the administration of the agent accordingly.

In another aspect, the invention includes polynucleotides encoding a CPP of the invention, polynucleotides encoding a polypeptide having an amino acid sequence selected from one of the groups consisting of SEQ ID NOs: 1-5, 6-10, 11-14, 15-23 and 24-28, oligonucleotides complementary to CPP gene sequences for diagnostic and analytical assays (e.g., PCR, hybridization-based techniques), and vectors for expressing CPPs.

In another aspect, the invention provides a vector comprising DNA encoding a CPP. The invention also includes host cells and transgenic non-human animals comprising such a vector. There is also provided a method of making a CPP or CPP precursor. One preferred method comprises the steps of (a) providing a host cell containing an expression vector as disclosed above; (b) culturing the host cell under conditions whereby the DNA segment is expressed; and (c) recovering the protein encoded by the DNA segment. Another preferred method comprises the steps of: (a) providing a host cell capable of expressing a CPP; (b) culturing said host cell under conditions that allow expression of

said CPP; and (c) recovering said CPP. Within one embodiment the expression vector further comprises a secretory signal sequence operably linked to the DNA segment, the cell secretes the protein into a culture medium, and the protein is recovered from the medium. An especially preferred method of making a CPP includes chemical synthesis using standard peptide synthesis techniques, as described in the section titled "Chemical Manufacture of CPP Compositions" and in Example 2.

In another aspect, the invention includes isolated antibodies specific for any of the polypeptides, peptide fragments, or peptides described above. Preferably, the antibodies of the invention are monoclonal antibodies. Further preferred are antibodies that bind to a CPP exclusively, that is, antibodies that do not recognize other polypeptides with high affinity. Anti-CPP antibodies have purification, diagnostic and prognostic applications. Preferred anti-CPP antibodies for purification and diagnosis are attached to a label group. Preferred CPP-related disorders for diagnosis include coronary artery disease (CAD), coronary heart disease (CHD), peripheral vascular disease, cerebral ischemia (stroke), congestive heart failure, atherosclerosis, hypertension, and other cardiovascular diseases. Diagnostic methods include, but are not limited to, those that employ antibodies or antibody-derived compositions specific for a CPP antigen. Diagnostic methods for detecting CPPs in specific tissue samples and biological fluids (preferably plasma), and for detecting levels of expression of CPPs in tissues, also form part of the invention. Compositions comprising

within the scope of the invention, for example, for in vivo diagnosis and drug screening assays.

The invention further provides methods for diagnosis of cardiovascular disorders that comprise detecting in a sample of body fluid, preferably blood plasma, the presence or level of at least one CPP disclosed herein or any combination thereof. Further included are methods of using CPP compositions, including primers complementary to CPP genes and/or messenger RNA and anti-CPP antibodies, for detecting and measuring quantities of CPPs in tissues and biological fluids, preferably plasma. These methods are also suitable for clinical screening, prognosis, monitoring the results of therapy, identifying patients most likely to respond to a particular therapeutic treatment, drug screening and development, and identifying new targets for drug treatment.

The invention provides kits that may be used in the above-recited methods and that may comprise single or multiple preparations, or antibodies, together with other reagents, label groups, substrates, if needed, and directions for use. The kits may be used for diagnosis of disease, or may be assays for the identification of new diagnostic and/or therapeutic agents.

In one embodiment, Coronary Artery Disease (CAD) is defined by the appearance of at least one symptom. Such symptoms become more serious as the disease progresses. CAD is often

accompanied by reduced left ventricle capacity or output. Early CAD symptoms include elevated plasma levels of cholesterol and low-density lipoprotein (especially oxidized forms), as well as platelet-rich plasma aggregations. The vascular endothelium responds to inflammation and thus formation of plaques and levels of inflammatory and fibrinogenic factors increase. In addition, CAD, or atherosclerosis, is characterized by vascular calcification and hardening of the arteries. The resulting partial occlusion of the blood vessels leads to hypertension and ischemic heart disease. Eventual complete vascular occlusion results in myocardial infarction, stroke, or gangrene.

In a preferred embodiment, detection of increased plasma levels of at least one CPP of the invention indicates an increased risk that an individual will develop CAD. Preferably, said detection indicates that an individual has at least a 1.05-fold, 1.1-fold, 1.15-fold, and more preferably at least a 1.2-fold increased likelihood of developing CAD. Alternatively, detection of increased plasma levels of at least one CPP of the invention indicates that an individual has CAD. The amount of CPP increase observed in an individual compared to a control sample will correlate with the certainty of the prediction or diagnosis of CAD. As individual plasma CPP levels will vary depending on family history and other risk factors, each will preferably be examined on a case-by-case basis. In preferred embodiments, CPP is detected in a human plasma sample by the methods of the invention. Especially preferred techniques are mass spectrometry and immunodetection. Preferably, a prediction or

in the experimental CPP level as compared to the control.

The invention further includes methods of using CPP-modulating compositions to prevent or treat disorders associated with aberrant expression or processing of CPPs of SEQ ID NOs:1-28 in an individual. Preferred CPP-related disorders include coronary artery disease (CAD), coronary heart disease (CHD), peripheral vascular disease, cerebral ischemia (stroke), congestive heart failure, atherosclerosis, hypertension, and other cardiovascular diseases. A preferred embodiment of the invention is a method of preventing or treating a CPP-related disorder in an individual comprising the steps of: determining that an individual suffers from or is at risk of a CPP-related disorder and introducing a CPP-modulating composition to said individual.

Further aspects of the invention are also described in the specification and in the claims.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

SEQ ID NOs:1 and 2 describe the amino acid sequences of the polypeptides present in plasma samples of individuals with coronary artery disease (hereinafter, SEQ ID NO:2 is designated CPP 2).

SEQ ID NOs:3-5 are the amino acid sequences of tryptic peptides found in MS-MS and/or MS-MALDI mass spectrometry in plasma samples of individuals with coronary artery disease.

SEQ ID NO:6 describes the amino acid sequence of eosinophil-derived neurotoxin (EDN, hereinafter CPP 9), whereas SEQ ID NO:7 is the polypeptide sequence of the mature protein.

SEQ ID NOs:8, 9, and 10 are the amino acid sequences of tryptic peptides found by MS-MS mass spectrometry in plasma samples of individuals with coronary artery disease.

SEQ ID NO:11 describes the amino acid sequence of Human Epididymal secretory protein (HE) 1, whereas SEQ ID NO:12 is the polypeptide sequence of the mature protein (hereinafter, CPP 21).

SEQ ID NOs:13-14 are the amino acid sequences of tryptic peptides found by MS-MS mass spectrometry at a higher level in plasma samples of individuals with coronary artery disease.

SEQ ID NO:15 describes the amino acid sequence of Defensin 1 precursor, whereas SEQ ID NO:16 is the polypeptide sequence of the preprotein and SEQ ID NO:17 is the sequence of Defensin 1 (hereinafter, CPP 17).

SEQ ID NOs:18-23 are the amino acid sequences of tryptic peptides found by MS-MS mass spectrometry predominantly in plasma samples of individuals with coronary artery disease.

SEQ ID NO:24 describes the amino acid sequence of Plasminogen-related protein B

CPP 20).

SEQ ID NOs:26-28 are the amino acid sequences of tryptic peptides found by tandem mass spectrometry in plasma samples of individuals with Coronary Artery Disease.

BRIEF DESCRIPTION OF THE FIGURE

Figure 1 shows mature human Colipase polypeptide sequences (SEQ ID NOs:1 and 2) and the sequences of the tryptic peptides identified by tandem mass spectrometry in the plasma of individuals with coronary artery disease (SEQ ID NOs:3-5). The tryptic peptides are underlined in SEQ ID NOs:1 and 2.

Figure 2 shows the sequence of CPP 9 (SEQ ID NO:6) and the peptide sequences found by MS-MS mass spectrometry in the plasma of individuals with coronary artery disease (SEQ ID NOs:8-10). The tryptic peptides observed by tandem mass spectrometry are underlined in SEQ ID NOs:6 and 7. The signal peptide is double-underlined.

Figure 3 shows the sequence of the precursor protein (SEQ ID NO:11) and of the mature protein (SEQ ID NO:12, CPP 21) and peptide sequences found by MS-MS mass spectrometry in the

plasma of individuals with coronary artery disease (SEQ ID NOs:13-14). The tryptic peptides observed by tandem mass spectrometry are underlined in SEQ ID NOs:11 and 12. The signal peptide is double-underlined.

Figure 4 shows the sequence of CPP 17 (SEQ ID NO:17) and the peptide sequences found by MS-MS mass spectrometry in the plasma of individuals with coronary artery disease (SEQ ID NOs:18-23). The tryptic peptides observed by tandem mass spectrometry are underlined in SEQ ID NOs:15 and 16, which represent the precursor and preprotein, respectively. The signal peptide is double-underlined in SEQ ID NO:15.

Figure 5 shows the sequence of the precursor (SEQ ID NO:24) and mature protein (CPP20, SEQ ID NO:25) of the invention, and peptide sequences found by tandem mass spectrometry in the plasma of individuals with Coronary Artery Disease (SEQ ID NOs:26-28). The tryptic peptides observed by tandem mass spectrometry are underlined in SEQ ID NOs:24 and 25. The signal peptide is double-underlined in SEQ ID NO:24.

DETAILED DESCRIPTION OF THE INVENTION

The present invention described in detail below provides methods, compositions, and kits useful for screening, diagnosis, and prognosis of a cardiovascular disorder in a mammalian individual; the results of cardiovascular disorder therapy; for screening CPP modulators; and for drug development. The invention also encompasses the administration of therapeutic compositions to a mammalian individual to treat or prevent cardiovascular disorders. The mammalian individual may be a non-human mammal, but is preferably human, more preferably a human adult. For clarity of disclosure, and not by way of limitation, the invention will be described with respect to the analysis of blood plasma samples. However, as one skilled in the art will appreciate, the assays and techniques described below can be applied to other biological fluid samples (e.g. cerebrospinal fluid, lymph, bile, serum, saliva or urine) or tissue samples from an individual at risk of having or developing a cardiovascular disorder. The methods and compositions of the present invention are useful for screening, diagnosis and prognosis of a living individual, but may also be used for postmortem diagnosis in an individual, for example, to identify family members who are at risk of developing the same disorder.

Definitions

As used herein, the term "nucleic acids" and "nucleic acid molecule" is intended to include

DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule. Also, used interchangeably herein are terms "nucleic acids", "oligonucleotides", and "polynucleotides".

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated CPP nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free

nucleic acid as a hybridization probe, CPP nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning. A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

As used herein, the term "hybridizes to" is intended to describe conditions for moderate stringency or high stringency hybridization, preferably where the hybridization and washing conditions permit nucleotide sequences at least 60% homologous to each other to remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85%, 90%, 95% or 98% homologous to each other typically remain hybridized to each other. Stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. In a preferred, non-limiting example, stringent hybridization conditions for nucleic acid interactions are as follows: the hybridization step is realized at 65°C in the presence of 6 x SSC buffer, 5 x Denhardt's solution, 0,5% SDS and 100µg/ml of salmon sperm DNA. The hybridization step is followed by four washing steps:

- two washings during 5 min, preferably at 65°C in a 2 x SSC and 0.1%SDS buffer;
- one washing during 30 min, preferably at 65°C in a 2 x SSC and 0.1% SDS buffer,
- one washing during 10 min, preferably at 65°C in a 0.1 x SSC and 0.1%SDS buffer,

these hybridization conditions being suitable for a nucleic acid molecule of about 20 nucleotides in length. It will be appreciated that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art, for example be adapted according to the teachings disclosed in Hames B.D. and Higgins S.J. (1985) Nucleic Acid Hybridization: A Practical Approach. Hames and Higgins Ed., IRL Press, Oxford; and Current Protocols in Molecular Biology.

"Percent homology" is used herein to refer to both nucleic acid sequences and amino acid sequences. Amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology". To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably

nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position. The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology=# of identical positions/total # of positions 100).

The comparison of sequences and determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Karlin and Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264-68, modified as in Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-77, the disclosures of which are incorporated herein by reference in their entireties. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to the sequences of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to the

polypeptide sequences of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Research 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov, the disclosures of which are incorporated herein by reference in their entireties. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989), the disclosures of which are incorporated herein by reference in their entireties. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-translational modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl, acetyl, phosphate, amide, lipid, carboxyl, acyl, or carbohydrate groups are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino

naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "protein" as used herein may be used synonymously with the term "polypeptide" or may refer to, in addition, a complex of two or more polypeptides which may be linked by bonds other than peptide bonds, for example, such polypeptides making up the protein may be linked by disulfide bonds. The term "protein" may also comprehend a family of polypeptides having identical amino acid sequences but different post-translational modifications, particularly as may be added when such proteins are expressed in eukaryotic hosts.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein of the invention (i.e., CPP or biologically active fragment thereof) is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of a protein according to the invention in which the protein is separated from cellular components of the cells from which it is isolated or

recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of a protein according to the invention having less than about 30% (by dry weight) of protein other than the protein of the invention (also referred to herein as a "contaminating protein"), more preferably less than about 20% of protein other than the protein according to the invention, still more preferably less than about 10% of protein other than the protein according to the invention, and most preferably less than about 5% of protein other than the protein according to the invention. When the protein according to the invention or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of a protein of the invention in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of a protein of the invention having less than about 30% (by dry weight) of chemical precursors or non-protein chemicals, more preferably less than about 20% chemical precursors or non-protein chemicals, still more preferably less than about 10% chemical precursors or non-protein chemicals,

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

The term "Cardiovascular disorder Plasma Polypeptide" or "CPP" refers to a polypeptide comprising the sequence described by any one of SEQ ID NOs:1-28. Such polypeptide may be post-translationally modified as described herein. CPPs may also contain other structural or chemical modifications such as disulfide linkages or amino acid side chain interactions such as hydrogen and amide bonds that result in complex secondary or tertiary structures. CPPs also include mutant polypeptides, such as deletion, addition, swap, or truncation mutants, fusion polypeptides comprising such polypeptides, and polypeptide fragments of at least three, but preferably, and where applicable, 8, 10, 12, 15, or 21 contiguous amino acids of the sequence of SEQ ID NOs:1-28. Further included are CPP proteolytic precursors and intermediates of the sequence selected from the group consisting of SEQ ID NOs:1-28. The invention embodies polypeptides encoded by the nucleic acid sequences of CPP genes or CPP mRNA species, preferably human CPP genes and mRNA species, including

isolated CPPs consisting of, consisting essentially of, or comprising the sequence of SEQ ID NOs:1-28. Preferred CPPs have a sequence comprising one of the sequences of SEQ ID NOs: 1-2, 6-7, 11-12, 15-17, and 24-25. Preferred CPP fragments have a sequence comprising one of the sequences of SEQ ID Nos: 3-5, 8-10, 13-14, 18-23 and 26-28. Preferred CPPs retain at least one biological activity of CPPs of SEQ ID NOs:1-28.

The term "biological activity" as used herein refers to any single function carried out by a CPP. These include but are not limited to: (1) indicating that an individual has or will have a cardiovascular disorder; (2) circulating through the bloodstream of individuals with a cardiovascular disorder; (3) antigenicity, or the ability to bind an anti-CPP specific antibody; (4) immunogenicity, or the ability to generate an anti-CPP specific antibody; and for CPP 2: (5) interacting with a CPP target protein, preferably a lipase; (6) stabilizing the active site of a lipase; (7) increasing lipase activity; (8) interacting with a CPP target molecule such as a phospholipid, micelle, or triglyceride; and (9) forming at least one disulfide bond; for CPP 9: (5) forming intramolecular amino acid side chain interactions such as hydrogen, amide, or preferably disulfide links; (6) interaction with a CPP target molecule, preferably an RNA molecule or virion (such as respiratory syncytial virus or RSV); (7) antiviral activity, and (8) hydrolysis of RNA phosphodiester bonds; for CPP 21: (5) forming intramolecular amino acid side chain interactions such as hydrogen, amide, or preferably disulfide

forming intramolecular amino acid side chain interactions such as hydrogen, amide, or preferably disulfide links; (6) interaction with a CPP target molecule, preferably a bacterial endotoxin; (7) neutralizing bacterial endotoxins; (8) promoting mast cell chemotaxis; (9) undergoing posttranslational processing, for example, specific proteolysis; (10) functioning as an antimicrobial defense; and (11) inhibiting contraction of smooth muscle cells; and for CPP 20: (5) forming intramolecular amino acid side chain interactions such as hydrogen, amide, or preferably disulfide links; (6) interacting with a CPP target molecule, in particular, a kringle domain-containing peptides such as plasminogen; and (7) reducing tumor growth.

As used herein, a "CPP modulator" is a molecule (e.g., polynucleotide, polypeptide, small molecule, or antibody) that is capable of modulating (i.e., increasing or decreasing) either the expression or the biological activity of the CPPs of the invention. A CPP modulator that enhances CPP expression or activity is described as a CPP activator or agonist. Conversely, a CPP modulator that represses CPP expression or activity is described as a CPP inhibitor or antagonist. Preferably, CPP modulators increase/ decrease the expression or activity by at least 5, 10, or 20%. CPP inhibitors include anti-CPP antibodies, fragments thereof, antisense polynucleotides, and molecules

characterized by screening assays, as described herein. CPP agonists include polynucleotide expression vectors and molecules characterized by screening assays as described herein.

A "CPP-related disorder" or "CPP-related disease" describes a cardiovascular disorder. Preferred disorders include coronary artery disease (CAD), coronary heart disease (CHD), peripheral vascular disease, cerebral ischemia (stroke), congestive heart failure, atherosclerosis, hypertension, and other cardiovascular diseases. Preferably, the likelihood that an individual will develop or already has such a disorder is indicated by higher than normal plasma levels of at least one CPP.

Another aspect of the invention pertains to anti-CPP antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen-binding site which specifically binds (immunoreacts with) an antigen, such as CPP, or a biologically active fragment or homologue thereof. Preferred antibodies bind to a CPP exclusively and do not recognize other polypeptides with high affinity. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind a CPP, or a biologically active fragment or homologue thereof. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one

monoclonal antibody composition thus typically displays a single binding affinity for a particular CPP with which it immunoreacts. Preferred CPP antibodies are attached to a label group.

As used herein, a "label group" is any compound that, when attached to a polynucleotide or polypeptide (including antibodies), allows for detection or purification of said polynucleotide or polypeptide. Label groups may be detected or purified directly or indirectly by a secondary compound, including an antibody specific for said label group. Useful label groups include radioisotopes (e.g., ^{32}P, ^{35}S, ^{3}H, ^{125}I), fluorescent compounds (e.g., 5-bromodesoxyuridin, umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride, phycoerythrin acetylaminofluorene, digoxigenin), luminescent compounds (e.g., luminol, GFP, luciferin, aequorin), enzymes or enzyme co-factor detectable labels (e.g., peroxidase, luciferase, alkaline phosphatase, galactosidase, or acetylcholinesterase), or compounds that are recognized by a secondary factor such as strepavidin, GST, or biotin. Preferably, a label group is attached to a polynucleotide or polypeptide in such a way as to not interfere with the biological activity of the polynucleotide or polypeptide.

Radioisotopes may be detected by direct counting of radioemission, film exposure, or by

scintillation counting, for example. Enzymatic labels may be detected by determination of conversion of an appropriate substrate to product, usually causing a fluorescent reaction. Fluorescent and luminescent compounds and reactions may be detected by, e.g., radioemission, fluorescent microscopy, fluorescent activated cell sorting, or a luminometer.

As used herein with respect to antibodies, an antibody is said to "selectively bind" or "specifically bind" to a target if the antibody recognizes and binds the target of interest but does not substantially recognize and bind other molecules in a sample, e.g., a biological sample, which includes the target of interest.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked.

recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

As used herein, "effective amount" describes the amount of an agent, preferably a CPP modulator of the invention, sufficient to have a desired effect. For example, an anticardiovascular disorder effective amount is the amount of an agent required to reduce a symptom of a cardiovascular disorder in an individual by at least 1, 2, 5, 10, 15, or preferably 25%. The term may also describe the amount of an agent required to ameliorate a cardiovascular disorder-caused symptom in an individual. Common symptoms of cardiovascular disorders include: chest pressure, heartburn, nausea, vomiting, numbness, shortness of breath, heavy cold sweating, unexplained fatigue, and feelings of anxiety. The more severe symptoms of cardiovascular disorders are chest pain (angina pectoris), rhythm disturbances (arrhythmias), stroke, or heart attack. The effective amount for a particular patient may vary depending on such factors as the diagnostic method of the symptom being measured, the state of

the condition being treated, the overall health of the patient, method of administration, and the severity of side-effects.

CPPs of the invention

The Cardiovascular disorder Plasma Polypeptides (CPPs) of the invention are described in the sequence listing as SEQ ID NOs: 1-2, 6-7, 11-12, 15-17, and 24-25. CPPs comprising an amino acid sequence selected from one of the groups consisting of SEQ ID NOs: 1-2, 6-7, 11-12, 15-17, and 24-25 are secreted and circulate at a higher level in blood plasma of individuals that have or are at risk of developing a cardiovascular disorder.

Further included CPPs are polypeptides comprising an amino acid sequence selected from one of the groups consisting of SEQ ID NOs: 3-5, 8-10, 13-14, 18-23 and 26-28. Such CPPs also are secreted and circulate in the plasma. Preferably, such CPPs also comprise additional amino acids from one of the groups of SEQ ID NO: 1-2, 6-7, 11-12, 15-17, and 24-25. Such additional amino acids are fused in frame with the selected sequence to form contiguous amino acid sequence from the proteins of SEQ ID NO: 1-2, 6-7, 11-12, 15-17, and 24-25.

Interestingly, levels of the CPPs of the invention are increased in the plasma of individuals suffering from cardiovascular disorders. As such, the CPPs of the invention provide a useful

presence of, a cardiovascular disorder. Further, CPPs are useful for drug design and in therapeutic strategies for prevention and treatment of cardiovascular disorders.

CPP 2, of SEQ ID NO:2, represents the sequence of the protein Colipase. Colipase is a small protein cofactor for pancreatic triglyceride lipase (PTL), which is required for efficient dietary lipid hydrolysis. Mature colipase is 90 amino acids in length, with five conserved disulfide bonds. The protein is synthesized as preprocolipase, with 112 amino acids, and processed to procolipase, from which an additional 5 amino acids are cleaved at the N-terminus. Colipase is a secreted protein that binds to the C-terminal, non-catalytic domain of the lipase, thereby contributing to the hydrophobic active binding site; stabilizing activity in the presence of inhibitory substances such as bile acids; and directing the enzyme to the oil-water interface. Five residues of colipase form polar interactions with PTL: Arg39, Glu40, Glu59, Arg60, and Asn84. In the active conformation, Glu10 and Arg33 also interact with PTL. The lipase-colipase complex is activated by micelles, which also stabilize the open conformation and expose the hydrophobic active site. PTL and colipase function in the duodenum, releasing 50-70% of dietary fatty acids (see Crandall WV and Lowe ME, J Biol Chem (2001)

276:12505-12; van Tilbeurgh H, et al., Biochim Biophys Acta (1999) 1441:173-84; Pignol D, et al., JBC, (2000) 275:4220-4224; and Lowe ME, et al., Annu. Rev. Nutr (1997) 17:141-58).

As colipase is normally present and active in the duodenum, its presence in human blood plasma is indicative of a biological abnormality. This is further evidenced by the absence of the CPPs of the present invention in normal human plasma. Colipase has previously been observed in the serum of patients with acute pancreatitis (Junge, W. and Leybold, K, Clin.Chim.Acta (1982) 123(3), 293-302). As such, CPPs of the invention are useful for drug design and in therapeutic strategies for prevention and treatment of cardiovascular disorders. Interestingly in this regard, a therapeutic approach against obesity has been developed through the use of a lipase inhibitor, Orlistat (Sternby, B. et al., Clin.Nutr. (2002) 21(5), 395-402). The inhibitor is thought to act through the reduction of triglycerides intake, and has been observed to cause steatorrhoea, a condition also associated with defects in colipase (Gaskin, K.J. et al., Gastroenterology (1984) 86(1), 1-7).

Full length CPP 9 (SEQ ID NO:7) forms four disulfide bonds; is glycosylated, including a rare C-linked glycosylation; is a major eosinophil granule component, degrades RNA phosphodiester bonds, and has antiviral activity. Full length CPP 9 has the sequence of nonsecretory RNase 2. Ribonucleases catalyze the hydrolysis of phosphodiester bonds in RNA chains. Human ribonucleases (EDN), nonsecretory RNase 3 or eosinophil cationic protein (ECP), RNase 4, angiogenin, and the recently described RNase k6. RNases 2-4 and 6, which belong to the RNase A family, have antipathogenic activities (Zhang J and Rosenberg HF, Genetics 2000; 156:1949-58). EDN, a nonsecretory ribonuclease, degrades mRNA by hydrolyzing single-stranded polyribonucleotides in the 3' to 5' direction. EDN is a major eosinophil granule protein, and as such is found in tissues affected by eosinophils, including pancreas, liver, lung, spleen, and body fluids.

Increased serum levels of human Rnase 1 are indicative of pancreatic cancer (Fernandez-Salas, et al., Eur. J. Biochem., 2000, 267(5):1484-94). Intraventricular injection of EDN, in particular, has been implicated in promoting the Gordon phenomenon, a hyper-eosinophilic syndrome resulting in loss of Purkinje cells and cerebellar dysfunction (Newton, D.L. et al., J.Neurosci., 1994, 14(2): 538-44). In addition, eosinophils and EDN respond to the respiratory disease caused by the single-stranded RNA virus respiratory syncytial virus (RSV; family Paramyxoviridae). Eosinophils are recruited to the lung parenchyma in response to RSV infection. Once recruited, they interact with and are activated directly by RSV virions, resulting in degranulation (Domachowske JB, et al., Nucleic Acids Res 1998 Dec 1;26(23):5327-32). EDN promotes the destruction of extracellular RSV virions

via an RNase-dependent, yet EDN-specific, mechanism.

The CPP 21 of the invention (SEQ ID NO:12) represent a plasma form of the Human Epididymal secretory protein (HE)1. HE1, a novel human epididymal gene product isolated by differential screening, predicts an abundant, small secretory glycopeptide. HE1 is encoded by a well-conserved, single-copy gene (Kirchhoff, et al., Biol Reprod, 1996, 54:847-56). The gene is broadly expressed, with highest levels in testies, kidney, and liver. HE1 protein is normally present in the lysosomal compartment and makes up a major component of epididymal fluid secretions.

The inherited neurodegenerative disease Niemann-Pick Type C2 (NPC2) results from mutations in the HE1 (or NPC2) gene that cause abnormally high cellular cholesterol accumulation (Naureckiene S et al., *Science* (2000) 290:2298 and WO 02059369). Niemann Pick type C disease is associated with intracellular cholesterol and glycolipid trafficking defects. (Ong YY et al., Exp Brain Res (2001) 141:218-31). In NPC2 patients, cholesterol is not transported through the late endosomal / lysosomal system and accumulates in the lysosomes, meanwhile depriving cells of free cholesterol. Consequently, the cells produce additional cholesterol to compensate for the defect, despite the already excessive cholesterol accumulation (Garver WS et al., *Curr Mol Med*, (2002) 2:485-505). A hydrophobic, sterol-binding pocket of HE1 specifically interacts with cholesterol with high affinity

Using the methods of the current invention, HE1 has been discovered in the plasma of Coronary Artery Disease patients. This protein is normally associated with intracellular trafficking and epididymal secretions, and thus is not expected in the plasma of control individuals. Thus, HE1 is an important biomarker for cardiovascular disorders.

CPP 17 (SEQ ID NO:17) is the plasma form of the antimicrobial peptides Defensin 1-3. Defensins 1-3 (alpha-defensins) were cloned from a leukemia cell library, and expression was found in normal bone marrow and circulating lymphocytes (Daher, et al., PNAS. 85:7327-7331 (1988)). Alpha-defensins are constitutively expressed in neutrophils and can make up to 5% of cellular protein. The full-length proteins are processed to form 29-35 amino acid peptides that are released in response to microbial and viral infection. These cationic peptides insert and disrupt prokaryotic membranes (Hill, et al. Science 251:1481-1485 (1991)). In addition, alpha-defensins reduce HIV replication in response to the CD8 T cell factor CAF (Zhang, et al. Science 298:995-1000 (2002)).

Alpha-defensins bind low-density lipoprotein (LDL) particles in the blood and prevent LDL receptor-binding and subsequent degradation (Higazi, et al. Blood 96:1393-1398 (2000) and

references cited therein). It has since been discovered that alpha-defensins interact with vesicular smooth muscle cells and inhibit contraction (Nassar, et al. Blood 100:4026-32 (2002)).

The inventors have revealed an elevated level of alpha-defensin peptides in the plasma of individuals with Coronary Artery Disease. These results provide direct evidence that alpha-defensins are useful plasma biomarkers for the disease.

CPP 20 (SEQ ID NO:25) is a plasma form of the Plasminogen-related Protein B. Plasminogen-related protein B (also Plasimilar) was cloned from chondrocytes and is highly homologous to the N terminus of plasminogen (Ichinose, Biochemistry 31:3113-8 (1992)). The 96 amino acid-long precursor protein is secreted after removal of a 19 amino acid signal sequence. Plasminogen-related protein B binds to kringle domains and is thought to interfere with plasminogen binding to fibrin or alpha-2 anti-plasmin (Lewis, et al., Eur J Biochem. 259:618-625 (1999)). Plasminogen-related protein B is expressed in metastatic tumor cells (WO 9321341and Weissbach and Treadwell, Biochem. Biophys. Res. Commun. 186:1108-1114 (1992)). In addition, Lewis, et al. (Anticancer Res 21:2287-91 (2001) demonstrated that the protein reduces the growth of tumors explanted in mice, while plasminogen has little effect. Thus, it is described as a treatment for malignant conditions as well as a treatment for angiogenesis-related disorders (WO 9946282).

discovered in the plasma of Coronary Artery Disease patients. This protein is usually associated with the extracelluar matrix and is undetectable in the plasma of control individuals. Thus, Plasminogen-related protein B is an important biomarker for cardiovascular disorders.

The terms "Cardiovascular disorder Plasma Polypeptide" and "CPP" are used herein to embrace any and all of the peptides, polypeptides and proteins of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies CPPs from humans, including isolated or purified CPPs consisting of, consisting essentially of, or comprising an amino acid sequence selected from one of the groups consisting of SEQ ID NOs: 1-5, 6-10, 11-14, 15-23 and 24-28. Further included are unmodified precursors, proteolytic precursors and intermediates of the sequence selected from one of the groups consisting of SEQ ID NOs: 1-5, 6-10, 11-14, 15-23 and 24-28.

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 3 amino acids, preferably at least 8 to 10 amino acids, with a CPP

biological activity. In preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the CPP sequence. The invention also concerns the polypeptide encoded by the CPP nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof.

One aspect of the invention pertains to isolated CPPs, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-CPP antibodies. In one embodiment, native CPP peptides can be isolated from plasma, cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, CPPs are produced by recombinant DNA techniques. Alternative to recombinant expression, a CPP can be synthesized chemically using peptide synthesis techniques, as described in the section titled "Chemical Manufacture of CPP compositions" and in Example 2.

Typically, biologically active portions comprise a domain or motif with at least one activity of a CPP. A biologically active CPP may, for example, comprise at least 1, 2, 3, or 5 amino acid changes from the sequence selected from one of the groups consisting of SEQ ID NOs: 1-5, 6-10, 11-14, 15-23 and 24-28, or comprise at least 1%, 2%, 3%, 5%, 8%, 10% or 15% change in amino acids from the sequence selected from one of the groups consisting of SEQ ID NOs: 1-5, 6-10, 11-14, 15-23 and 24-28.

Characterization of CPPs

The polypeptides of the invention, CPPs, are defined by the tryptic peptides of SEQ ID NOs: 3-5, 8-10, 13-14, 18-23 and 26-28, listed in Table 1. These peptides were isolated from the plasma of Coronary Artery Disease patients and characterized according to the MicroProt.TM method, as described in Example 1. SEQ ID NOs: 1-2, 6-7, 11-12, 15-17, and 24-25 represent the polypeptide species found in CAD plasma from which the tryptic peptides were released.

The CPPs of the invention are all less than or around 20kD in molecular weight, as the plasma sample is first separated based on molecular weight. Higher molecular weight polypeptide species are separated and characterized by a different method. As described in Example 1, the plasma sample is subjected to a number of chromatography separations. Details about these chromatography methods are given in Example 1.

The first separation is on a cation exchange chromatography column, which is eluted with increasing salt concentration. Eighteen fractions are collected. The CEX column in Table 1 lists which fractions contained each tryptic peptide, as well as its elution conditions. Separation by cation exchange provides an indication of the overall positive charge of a polypeptide species. Cation

exchange is followed by a reverse phase HPLC separation. The RP1 column in Table 1 lists in which of the 30 fractions each tryptic peptide eluted, as well as its elution conditions. Separation by reverse phase provides an indication of the overall hydrophobicity of a polypeptide species. The last two digits of the column labeled Run Number indicate which of the 24 eluted fractions contained the tryptic peptide from the second reverse phase HPLC separation (see Example 1). Proteome indicates, for CPP 17, in which of the sample each tryptic peptide was observed. For CPP2, CPP 9, CPP 20 and CPP 21, all the observations were made solely in the Diseased (CAD) sample. Olav scores, when present, reflect, among other things, the strength of the experimental MS-MS signal over noise, as detected by the MS-MS data identification software, and thus give an indication of the protein concentration in the sample.

Table 1a – CPP 2

CPP #	Sequence	CEX	Salt	RP1	% B	Run
CPP 2	SNCCQHSSALGLAR	10	175 mM	9	35.7	89870_13
	SNCCQHSSALGLAR	10	175 mM	9	35.7	89870_14
	CTSMASENSECSVK	10	175 mM	9	35.7	89870_13
	CTSMASENSECSVK	10	175 mM	9	35.7	89870_14
	TIVGSITNTNEGICHDAGR	9	175 mM	9	35.7	108111_10

Table 1b – CPP 9

CPP #	Sequence	CEX	Salt	RP1	% B	Run
CPP 9 Cluster 1	DPPQYPVVPVHLDR	2	75 mM	8	33.7	84158_22
	DPPQYPVVPVHLDR	2	75 mM	8	33.7	84158_22
	DPPQYPVVPVHLDR	2	75 mM	8	33.7	84158_22
	DPPQYPVVPVHLDR	2	75 mM	8	33.7	84158_22
	DPPQYPVVPVHLDR	2	75 mM	8	33.7	84158_22
	RDPPQYPVVPVHLDR	2	75 mM	8	33.7	84158_22
	RDPPQYPVVPVHLDR	2	75 mM	8	33.7	84158_23
	YAQTPANMFYIVACDNR	2	75 mM	8	33.7	84158_19
	YAQTPANMFYIVACDNR	2	75 mM	8	33.7	84158_20
	YAQTPANMFYIVACDNR	2	75 mM	8	33.7	84158_21
	YAQTPANMFYIVACDNR	2	75 mM	8	33.7	84158_22
	YAQTPANMFYIVACDNR	2	75 mM	8	33.7	84158_23
	YAQTPANMFYIVACDNR	2	75 mM	8	33.7	84158_24

CPP 9 Cluster 2	YAQTPANMFYIVACDNR	3	75 mM	10	37.5	84274_13
	YAQTPANMFYIVACDNR	3	75 mM	11	39.4	84286_14
	YAQTPANMFYIVACDNR	3	75 mM	11	39.4	84286_15
	RDPPQYPVVPVHLDR	4	75 mM	10	37.5	88113_12
	YAQTPANMFYIVACDNR	4	75 mM	10	37.5	88113_12
	YAQTPANMFYIVACDNR	4	75 mM	10	37.5	88113_13
	YAQTPANMFYIVACDNR	5	75 mM	11	39.4	101105_13
CPP 9 Cluster 3	YAQTPANMFYIVACDNR	8	100 mM	12	41.3	95713_07
	YAQTPANMFYIVACDNR	9	175 mM	11	39.4	105000_10
	YAQTPANMFYIVACDNR	9	175 mM	11	39.4	105000_13
	RDPPQYPVVPVHLDR	9	175 mM	12	41.3	105012_08
	YAQTPANMFYIVACDNR	9	175 mM	12	41.3	105012_08
	YAQTPANMFYIVACDNR	9	175 mM	12	41.3	105012_09

For CPP 9, the information provided in Table 1b reveals a number of characteristics of the polypeptide species present in plasma of CAD patients. The following represent exemplary interpretations of the data contained in Table 1b. Table 1b shows that the same tryptic peptide is found in more than one fraction of a given separation. Some of the fractions are widely spaced, which indicates that the tryptic peptide is released from different circulating polypeptide species. For example, CPP 9 was observed in a continuous set of CEX fractions from 2 to 5, and then again in another continuous set of CEX fractions from 8 to 9. This indicates the existence of at least two polypeptide species from which the tryptic peptides were derived, which possess different positive charge properties. This may indicate the presence of additional amino acids or a charged modification (e.g., phosphate group) in one species and not the other. Similarly, CPP 9 was observed in RP1 fraction 8, and then in a continuous set of RP1 fractions from 10 to 12. Once again, this indicates the existence of at least two polypeptides species from which the tryptic peptides were derived, which possess different hydrophobic indexes. Taken together, these observations indicate that the observed peptides fall into three non-sequence-specific elution clusters: the first elution cluster includes fragments eluted in CEX fraction 2 and RP1 fraction 8; the second cluster includes fragments eluting in CEX fractions 3-5 and RP1 fractions 10-11; and the third includes fragments eluting in CEX fractions 8-9 and RP1 fractions 11-12. Based on the combined fraction information for the tryptic peptides of SEQ ID NOs:3-5, it is likely that three different forms or species of CPP 9 circulate in plasma of individuals with CAD. These CPP 9 forms likely differ in post-translational modification and/or amino acid length. The tryptic peptides of SEQ ID NOs:8-10 are derived from the polypeptide

of SEQ ID NO:7, which is processed to form the different polypeptide species present in the plasma of individuals with CAD.

Table 1c – CPP 21

CPP #	Peptide Sequence	CEX	Salt	RP1	% B	Run Number
CPP 21	AVVHGILMGVPVPFPIPEPDGCK	3	75	13	43.27	84318_14
	AVVHGILMGVPVPFPIPEPDGCK	3	75	13	43.27	84318_15
	AVVHGILMGVPVPFPIPEPDGCK	3	75	13	43.27	84318_15
	AVVHGILMGVPVPFPIPEPDGCK	2	75	10	37.5	84178_22
	AVVHGILMGVPVPFPIPEPDGCK	3	75	13	43.27	84318_15
	SGINCPIQK	3	75	13	43.27	84318_15
	SGINCPIQK	2	75	10	37.5	84178_22

Table 1d – CPP 17

Table 1d – CPP 17

Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
YGTCIYQGR	6	100	8	33.65	130804_09	Control	12.59
YGTCIYQGR	6	100	8	33.65	130804_10	Control	20.62
YGTCIYQGR	6	100	9	35.7	130820_10	Control	30.84
IPACIAGER	7	100	7	31.73	130782_15	Control	23.41
IPACIAGER	7	100	7	31.73	130782_13	Control	16
IPACIAGER	7	100	7	31.73	130782_13	Control	12.8
IPACIAGER	7	100	7	31.73	130782_15	Control	11.73
IPACIAGER	7	100	7	31.73	130782_14	Control	23.61
YGTCIYQGR	7	100	7	31.73	130782_13	Control	44.04
YGTCIYQGR	7	100	7	31.73	130782_13	Control	52.39
YGTCIYQGR	7	100	7	31.73	130782_14	Control	49.92
YGTCIYQGR	7	100	7	31.73	130782_14	Control	47.62
YGTCIYQGR	7	100	7	31.73	130782_14	Control	36.78
YGTCIYQGR	7	100	7	31.73	130782_12	Control	42.07
YGTCIYQGR	7	100	7	31.73	130782_15	Control	43.83
IPACIAGER	7	100	8	33.65	130808_10	Control	27.5

Table 1d – CPP 17							
Peptide Sequence	**CEX**	**Salt**	**RP1**	**% B**	**Run**	**Proteome**	**Olav Score**
IPACIAGER	7	100	8	33.65	130808_08	Control	25.2
IPACIAGER	7	100	8	33.65	130808_11	Control	24.8
IPACIAGER	7	100	8	33.65	130808_09	Control	23.61
YGTCIYQGR	7	100	8	33.65	130808_08	Control	42.42
IPACIAGER	7	100	10	37.5	130836_07	Control	29.46
YGTCIYQGR	7	100	10	37.5	130836_07	Control	48.61
IPACIAGER	7	100	12	41.34	130868_12	Control	15.48
IPACIAGERR	8	100	8	33.65	121059_08	Control	24.97
IPACIAGER	8	100	8	33.65	121059_11	Control	21.05
RYGTCIYQGR	8	100	8	33.65	121059_08	Control	46.79
RYGTCIYQGR	8	100	8	33.65	121059_08	Control	35.9
YGTCIYQGR	8	100	8	33.65	121059_11	Control	28.98
IPACIAGER	8	100	9	35.7	121071_08	Control	12.3
IPACIAGER	8	100	9	35.7	121071_08	Control	28.34
IPACIAGER	8	100	9	35.7	121071_11	Control	28.63
IPACIAGER	8	100	9	35.7	121071_07	Control	12.19
IPACIAGER	8	100	9	35.7	121071_09	Control	33.23
YGTCIYQGR	8	100	9	35.7	121071_10	Control	34.47
YGTCIYQGR	8	100	9	35.7	121071_08	Control	41.24
YGTCIYQGR	8	100	9	35.7	121071_11	Control	52.02
YGTCIYQGR	8	100	9	35.7	121071_09	Control	53.73
YGTCIYQGR	8	100	9	35.7	121071_07	Control	35.05
YGTCIYQGR	8	100	9	35.7	121071_12	Control	25.16
LWAFCC	8	100	9	35.7	121071_09	Control	17.74
IPACIAGER	8	100	10	37.5	121091_08	Control	25.38
YGTCIYQGR	8	100	10	37.5	121091_07	Control	42.97
YGTCIYQGR	8	100	10	37.5	121091_08	Control	45.46
YGTCIYQGR	8	100	10	37.5	121091_09	Control	47.15
IPACIAGER	8	100	11	39.42	121095_07	Control	26.69
IPACIAGER	8	100	11	39.42	121095_08	Control	32.28
IPACIAGER	8	100	11	39.42	121095_09	Control	27.02
YGTCIYQGR	8	100	11	39.42	121095_09	Control	42.21
YGTCIYQGR	8	100	11	39.42	121095_14	Control	16.66
YGTCIYQGR	8	100	11	39.42	121095_09	Control	36.49

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
YGTCIYQGR	8	100	11	39.42	121095_07	Control	42.61
YGTCIYQGR	8	100	11	39.42	121095_08	Control	36.09
LWAFCC	8	100	11	39.42	121095_08	Control	17.02
LWAFCC	8	100	11	39.42	121095_07	Control	20.09
IPACIAGER	8	100	12	41.34	121115_12	Control	15.91
IPACIAGER	8	100	12	41.34	121115_05	Control	29.01
IPACIAGER	8	100	12	41.34	121115_07	Control	28.02
IPACIAGER	8	100	12	41.34	121115_06	Control	30.2
YGTCIYQGR	8	100	12	41.34	121115_09	Control	27
YGTCIYQGR	8	100	12	41.34	121115_07	Control	48.94
YGTCIYQGR	8	100	12	41.34	121115_06	Control	37.44
YGTCIYQGR	8	100	12	41.34	121115_05	Control	57.97
IPACIAGER	8	100	13	43.27	121127_06	Control	30.44
IPACIAGER	8	100	13	43.27	121127_05	Control	29.38
IPACIAGER	8	100	13	43.27	121127_07	Control	24.45
YGTCIYQGR	8	100	13	43.27	121127_06	Control	48.37
YGTCIYQGR	8	100	13	43.27	121127_05	Control	50.08
IPACIAGER	9	175	7	31.73	130774_13	Control	25.57
IPACIAGER	9	175	7	31.73	130774_11	Control	25.79
IPACIAGER	9	175	7	31.73	130774_13	Control	28.54
IPACIAGER	9	175	7	31.73	130774_12	Control	11.1
IPACIAGER	9	175	7	31.73	130774_12	Control	31.51
RYGTCIYQGR	9	175	7	31.73	130774_13	Control	18.08
YGTCIYQGR	9	175	7	31.73	130774_13	Control	48.73
YGTCIYQGR	9	175	7	31.73	130774_13	Control	48.83
YGTCIYQGR	9	175	7	31.73	130774_11	Control	42.52
YGTCIYQGR	9	175	7	31.73	130774_12	Control	61.96
YGTCIYQGR	9	175	7	31.73	130774_14	Control	60.25
YGTCIYQGR	9	175	7	31.73	130774_12	Control	44.13
YGTCIYQGR	9	175	7	31.73	130774_10	Control	14.34
LWAFCC	9	175	7	31.73	130774_12	Control	18.2
IPACIAGER	9	175	8	33.65	130786_11	Control	24.81
IPACIAGER	9	175	8	33.65	130786_08	Control	21.49
YGTCIYQGR	9	175	8	33.65	130786_08	Control	19.16

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
IPACIAGER	9	175	12	41.34	130860_10	Control	31.05
IPACIAGER	9	175	12	41.34	130860_11	Control	29.74
YGTCIYQGR	9	175	12	41.34	130860_10	Control	50.91
YGTCIYQGR	9	175	12	41.34	130860_11	Control	40.12
IPACIAGER	9	175	13	43.27	130876_10	Control	14.94
IPACIAGER	9	175	13	43.27	130876_12	Control	33.64
YGTCIYQGR	9	175	13	43.27	130876_12	Control	42.65
YGTCIYQGR	9	175	13	43.27	130876_10	Control	25.55
YGTCIYQGR	9	175	13	43.27	130876_12	Control	26.9
IPACIAGER	10	175	8	33.65	110332_08	Control	25.99
IPACIAGER	10	175	8	33.65	110332_11	Control	34.38
IPACIAGER	10	175	8	33.65	110332_12	Control	13.51
IPACIAGER	10	175	8	33.65	110332_09	Control	13.77
YGTCIYQGR	10	175	8	33.65	110332_10	Control	23.53
YGTCIYQGR	10	175	8	33.65	110332_11	Control	43.26
YGTCIYQGR	10	175	8	33.65	110332_08	Control	20.07
YGTCIYQGR	10	175	9	35.7	110348_09	Control	47.8
YGTCIYQGR	11	175	8	33.65	110336_12	Control	33.38
YGTCIYQGR	11	175	8	33.65	110336_11	Control	45.59
YGTCIYQGR	11	175	8	33.65	110336_12	Control	54.61
IPACIAGER	11	175	9	35.7	110344_10	Control	21.1
IPACIAGER	11	175	9	35.7	110344_09	Control	32.42
YGTCIYQGR	11	175	9	35.7	110344_09	Control	46.58
IPACIAGER	12	175	8	33.65	121083_08	Control	33.79
YGTCIYQGR	12	175	8	33.65	121083_08	Control	49.15
YGTCIYQGR	13	175	9	35.7	110352_09	Control	27.99
IPACIAGER	14	175	7	31.73	117862_13	Control	33.6
YGTCIYQGR	14	175	7	31.73	117862_13	Control	28.57
IPACIAGER	4	75	8	33.65	88177_09	CAD	15.34
IPACIAGER	5	75	9	35.7	101089_16	CAD	14.38
IPACIAGER	5	75	9	35.7	101089_11	CAD	18.62
YGTCIYQGR	5	75	9	35.7	101089_12	CAD	39.54
IPACIAGER	6	100	8	33.65	111981_09	CAD	26.89
IPACIAGER	6	100	9	35.7	100782_09	CAD	14.96

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
YGTCIYQGR	6	100	9	35.7	100782_10	CAD	12.18
YGTCIYQGR	6	100	9	35.7	100782_11	CAD	36.7
IPACIAGER	7	100	8	33.65	108099_11	CAD	26.83
IPACIAGER	7	100	8	33.65	108099_09	CAD	22.09
IPACIAGER	7	100	8	33.65	108099_12	CAD	13.52
IPACIAGER	7	100	8	33.65	108099_10	CAD	25.05
RYGTCIYQGR	7	100	8	33.65	108099_09	CAD	29.38
YGTCIYQGR	7	100	8	33.65	108099_08	CAD	49.46
YGTCIYQGR	7	100	8	33.65	108099_12	CAD	33.68
YGTCIYQGR	7	100	8	33.65	108099_09	CAD	46.69
YGTCIYQGR	7	100	8	33.65	108099_11	CAD	33.65
IPACIAGER	7	100	9	35.7	108103_11	CAD	18.78
IPACIAGER	7	100	9	35.7	108103_08	CAD	26.41
IPACIAGER	7	100	9	35.7	108103_10	CAD	28.48
IPACIAGER	7	100	9	35.7	108103_09	CAD	29.11
YGTCIYQGR	7	100	9	35.7	108103_09	CAD	47.95
YGTCIYQGR	7	100	9	35.7	108103_11	CAD	35.85
YGTCIYQGR	7	100	9	35.7	108103_10	CAD	51.77
IPACIAGER	7	100	10	37.5	105004_07	CAD	15.75
YGTCIYQGR	7	100	10	37.5	105004_08	CAD	43.36
YGTCIYQGR	7	100	10	37.5	105004_07	CAD	44.2
IPACIAGER	7	100	11	39.42	104992_08	CAD	31.77
YGTCIYQGR	7	100	11	39.42	104992_07	CAD	33.44
YGTCIYQGR	7	100	11	39.42	104992_08	CAD	28.08
IPACIAGER	7	100	12	41.34	105020_07	CAD	16.56
YGTCIYQGR	7	100	12	41.34	105020_07	CAD	13.61
IPACIAGER	8	100	8	33.65	98379_16	CAD	15.58
IPACIAGER	8	100	8	33.65	98379_11	CAD	25.88
IPACIAGER	8	100	8	33.65	98379_08	CAD	24.81
IPACIAGER	8	100	8	33.65	98379_12	CAD	26.89
RYGTCIYQGR	8	100	8	33.65	98379_09	CAD	38.01
RYGTCIYQGR	8	100	8	33.65	98379_09	CAD	35.93
YGTCIYQGR	8	100	8	33.65	98379_13	CAD	32.74
YGTCIYQGRLWAFCC	8	100	8	33.65	98379_08	CAD	49.34

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
YGTCIYQGR	8	100	8	33.65	98379_12	CAD	15.42
YGTCIYQGR	8	100	8	33.65	98379_09	CAD	50.78
YGTCIYQGR	8	100	8	33.65	98379_08	CAD	47.12
YGTCIYQGR	8	100	8	33.65	98379_08	CAD	46.31
YGTCIYQGR	8	100	8	33.65	98379_10	CAD	16.91
YGTCIYQGR	8	100	8	33.65	98379_10	CAD	36.03
YGTCIYQGR	8	100	8	33.65	98379_11	CAD	40.69
YGTCIYQGRLWAFCC	8	100	8	33.65	98379_08	CAD	41.85
YGTCIYQGR	8	100	8	33.65	98379_07	CAD	38.06
YGTCIYQGR	8	100	8	33.65	98379_12	CAD	52.75
LWAFCC	8	100	8	33.65	98379_11	CAD	17.02
LWAFCC	8	100	8	33.65	98379_08	CAD	20.09
LWAFCC	8	100	8	33.65	98379_09	CAD	20.09
LWAFCC	8	100	8	33.65	98379_12	CAD	21.28
IPACIAGER	8	100	9	35.7	98395_16	CAD	29.25
IPACIAGER	8	100	9	35.7	98395_12	CAD	21.9
IPACIAGER	8	100	9	35.7	98395_12	CAD	17.76
IPACIAGER	8	100	9	35.7	98395_12	CAD	20.93
IPACIAGER	8	100	9	35.7	98395_17	CAD	28.01
IPACIAGER	8	100	9	35.7	98395_09	CAD	23.18
IPACIAGER	8	100	9	35.7	98395_09	CAD	30.94
IPACIAGER	8	100	9	35.7	98395_08	CAD	20.52
IPACIAGER	8	100	9	35.7	98395_13	CAD	25.13
IPACIAGER	8	100	9	35.7	98395_12	CAD	30.76
IPACIAGERR	8	100	9	35.7	98395_11	CAD	26.5
IPACIAGER	8	100	9	35.7	98395_08	CAD	20.09
IPACIAGER	8	100	9	35.7	98395_09	CAD	16.16
IPACIAGER	8	100	9	35.7	98395_14	CAD	26.79
RYGTCIYQGR	8	100	9	35.7	98395_14	CAD	29.34
RYGTCIYQGR	8	100	9	35.7	98395_11	CAD	27.02
RYGTCIYQGR	8	100	9	35.7	98395_08	CAD	29.27
RYGTCIYQGR	8	100	9	35.7	98395_10	CAD	33.39
RYGTCIYQGR	8	100	9	35.7	98395_12	CAD	25.61
YGTCIYQGR	8	100	9	35.7	98395_14	CAD	41.77

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
YGTCIYQGRLWAFCC	8	100	9	35.7	98395_08	CAD	48.82
YGTCIYQGR	8	100	9	35.7	98395_12	CAD	51.26
YGTCIYQGRLWAFCC	8	100	9	35.7	98395_08	CAD	46.62
YGTCIYQGR	8	100	9	35.7	98395_11	CAD	38.68
YGTCIYQGR	8	100	9	35.7	98395_18	CAD	21.98
YGTCIYQGR	8	100	9	35.7	98395_13	CAD	38.58
YGTCIYQGR	8	100	9	35.7	98395_09	CAD	25.39
YGTCIYQGR	8	100	9	35.7	98395_17	CAD	36.39
YGTCIYQGR	8	100	9	35.7	98395_08	CAD	38.27
YGTCIYQGR	8	100	9	35.7	98395_08	CAD	43.61
YGTCIYQGRLWAFCC	8	100	9	35.7	98395_12	CAD	57.23
LWAFCC	8	100	9	35.7	98395_09	CAD	21.28
LWAFCC	8	100	9	35.7	98395_12	CAD	17.04
LWAFCC	8	100	9	35.7	98395_10	CAD	20.09
IPACIAGER	8	100	10	37.5	95685_11	CAD	27.27
IPACIAGER	8	100	10	37.5	95685_08	CAD	24.75
IPACIAGER	8	100	10	37.5	95685_10	CAD	16.84
IPACIAGER	8	100	10	37.5	95685_14	CAD	24.62
IPACIAGER	8	100	10	37.5	95685_09	CAD	19.98
IPACIAGER	8	100	10	37.5	95685_13	CAD	10.65
IPACIAGER	8	100	10	37.5	95685_06	CAD	31.36
RYGTCIYQGR	8	100	10	37.5	95685_06	CAD	31.07
RYGTCIYQGR	8	100	10	37.5	95685_07	CAD	38.34
YGTCIYQGR	8	100	10	37.5	95685_16	CAD	20.45
YGTCIYQGR	8	100	10	37.5	95685_14	CAD	32.12
YGTCIYQGR	8	100	10	37.5	95685_13	CAD	23.06
YGTCIYQGR	8	100	10	37.5	95685_07	CAD	38.95
YGTCIYQGR	8	100	10	37.5	95685_11	CAD	42.46
YGTCIYQGR	8	100	10	37.5	95685_15	CAD	19.41
YGTCIYQGR	8	100	10	37.5	95685_06	CAD	45.03
LWAFCC	8	100	10	37.5	95685_06	CAD	17.74
LWAFCC	8	100	10	37.5	95685_09	CAD	15.84
LWAFCC	8	100	10	37.5	95685_08	CAD	20.09
LWAFCC	8	100	10	37.5	95685_06	CAD	13.49

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
LWAFCC	8	100	10	37.5	95685_07	CAD	17.74
IPACIAGER	8	100	11	39.42	100948_09	CAD	25.95
IPACIAGER	8	100	11	39.42	100948_07	CAD	10.92
IPACIAGER	8	100	11	39.42	100948_11	CAD	13.34
IPACIAGER	8	100	11	39.42	100948_09	CAD	10.92
IPACIAGER	8	100	11	39.42	100948_05	CAD	12.53
IPACIAGER	8	100	11	39.42	100948_08	CAD	27.09
IPACIAGER	8	100	11	39.42	100948_06	CAD	27.26
IPACIAGER	8	100	11	39.42	100948_12	CAD	23.02
IPACIAGER	8	100	11	39.42	100948_07	CAD	17.37
IPACIAGER	8	100	11	39.42	100948_06	CAD	25.87
IPACIAGER	8	100	11	39.42	100948_07	CAD	27.12
RYGTCIYQGR	8	100	11	39.42	100948_06	CAD	23.52
YGTCIYQGR	8	100	11	39.42	100948_08	CAD	34.01
YGTCIYQGR	8	100	11	39.42	100948_09	CAD	29.05
YGTCIYQGR	8	100	11	39.42	100948_05	CAD	19.43
YGTCIYQGR	8	100	11	39.42	100948_06	CAD	39.1
YGTCIYQGR	8	100	11	39.42	100948_06	CAD	46.21
YGTCIYQGR	8	100	11	39.42	100948_06	CAD	42.1
LWAFCC	8	100	11	39.42	100948_06	CAD	14.67
LWAFCC	8	100	11	39.42	100948_07	CAD	21.28
LWAFCC	8	100	11	39.42	100948_07	CAD	20.09
IPACIAGER	8	100	12	41.34	95713_04	CAD	9.75
IPACIAGER	8	100	12	41.34	95713_04	CAD	16.21
IPACIAGER	8	100	12	41.34	95713_06	CAD	34.47
IPACIAGER	8	100	12	41.34	95713_04	CAD	20.55
IPACIAGER	8	100	12	41.34	95713_04	CAD	29.89
IPACIAGER	8	100	12	41.34	95713_09	CAD	28.65
IPACIAGER	8	100	12	41.34	95713_08	CAD	13.7
YGTCIYQGR	8	100	12	41.34	95713_08	CAD	12.11
YGTCIYQGR	8	100	12	41.34	95713_04	CAD	60.22
YGTCIYQGR	8	100	12	41.34	95713_06	CAD	55.12
IPACIAGER	8	100	13	43.27	95721_04	CAD	28.38
IPACIAGER	8	100	13	43.27	95721_05	CAD	28.84

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
IPACIAGER	8	100	13	43.27	95721_06	CAD	35.38
YGTCIYQGR	8	100	13	43.27	95721_04	CAD	58.93
YGTCIYQGR	8	100	13	43.27	95721_05	CAD	50.27
IPACIAGER	8	100	14	45.2	95729_02	CAD	18.23
IPACIAGER	8	100	14	45.2	95729_04	CAD	21.34
IPACIAGER	8	100	14	45.2	95729_03	CAD	31.93
YGTCIYQGR	8	100	14	45.2	95729_02	CAD	41.07
YGTCIYQGR	8	100	14	45.2	95729_04	CAD	36.34
LWAFCC	8	100	14	45.2	95729_03	CAD	9.71
ADEVAAAPEQIAADIP EVVVSLAWDESLAPK	8	100	14	45.2	95729_06	CAD	21.57
IPACIAGER	8	100	15	47.1	100961_03	CAD	14.11
YGTCIYQGR	8	100	15	47.1	100961_03	CAD	37.92
IPACIAGER	9	175	8	33.65	108107_11	CAD	22.35
IPACIAGER	9	175	8	33.65	108107_07	CAD	25.09
IPACIAGER	9	175	8	33.65	108107_08	CAD	27.93
IPACIAGER	9	175	8	33.65	108107_12	CAD	19.97
IPACIAGER	9	175	8	33.65	108107_09	CAD	24.95
RYGTCIYQGR	9	175	8	33.65	108107_09	CAD	41.98
RYGTCIYQGR	9	175	8	33.65	108107_11	CAD	33.57
RYGTCIYQGR	9	175	8	33.65	108107_08	CAD	44.88
RYGTCIYQGR	9	175	8	33.65	108107_11	CAD	25.03
IPACIAGER	9	175	9	35.7	108111_13	CAD	27.29
IPACIAGER	9	175	9	35.7	108111_20	CAD	21.68
IPACIAGER	9	175	9	35.7	108111_18	CAD	13.34
IPACIAGER	9	175	9	35.7	108111_14	CAD	29.32
IPACIAGER	9	175	9	35.7	108111_11	CAD	26.21
IPACIAGER	9	175	9	35.7	108111_09	CAD	28.44
IPACIAGER	9	175	9	35.7	108111_17	CAD	25.43
IPACIAGERR	9	175	9	35.7	108111_08	CAD	26.44
IPACIAGER	9	175	9	35.7	108111_15	CAD	26.58
IPACIAGER	9	175	9	35.7	108111_10	CAD	28.97
IPACIAGER	9	175	9	35.7	108111_08	CAD	25.32
IPACIAGER	9	175	9	35.7	108111_19	CAD	18.61

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
RYGTCIYQGR	9	175	9	35.7	108111_09	CAD	46.86
RYGTCIYQGR	9	175	9	35.7	108111_09	CAD	36.76
RYGTCIYQGR	9	175	9	35.7	108111_08	CAD	26.49
YGTCIYQGR	9	175	9	35.7	108111_13	CAD	18.83
YGTCIYQGR	9	175	9	35.7	108111_14	CAD	13.45
YGTCIYQGR	9	175	9	35.7	108111_11	CAD	17.98
LWAFCC	9	175	9	35.7	108111_09	CAD	14.67
LWAFCC	9	175	9	35.7	108111_09	CAD	17.76
LWAFCC	9	175	9	35.7	108111_08	CAD	16.32
LWAFCC	9	175	9	35.7	108111_11	CAD	14.67
IPACIAGER	9	175	10	37.5	104996_08	CAD	32.55
YGTCIYQGR	9	175	10	37.5	104996_07	CAD	46.07
YGTCIYQGR	9	175	10	37.5	104996_08	CAD	51.13
YGTCIYQGR	9	175	10	37.5	104996_08	CAD	46
LWAFCC	9	175	10	37.5	104996_07	CAD	13.49
LWAFCC	9	175	10	37.5	104996_07	CAD	15.84
IPACIAGER	9	175	11	39.42	105000_08	CAD	26.79
YGTCIYQGR	9	175	11	39.42	105000_08	CAD	31.81
LWAFCC	9	175	11	39.42	105000_08	CAD	18.91
IPACIAGER	9	175	12	41.34	105012_07	CAD	21.86
IPACIAGER	9	175	12	41.34	105012_07	CAD	22.47
YGTCIYQGR	9	175	12	41.34	105012_06	CAD	38.63
ADEVAAAPEQIAADIP EVVVSLAWDESLAPK	9	175	12	41.34	105012_08	CAD	73.1
IPACIAGER	9	175	14	45.2	105036_08	CAD	30.12
IPACIAGER	9	175	15	47.1	105028_09	CAD	31.09
IPACIAGER	9	175	15	47.1	105028_08	CAD	30.44
IPACIAGER	10	175	7	31.73	89838_15	CAD	12.17
IPACIAGER	10	175	7	31.73	89838_16	CAD	40.44
IPACIAGER	10	175	7	31.73	89838_21	CAD	25.98
YGTCIYQGR	10	175	7	31.73	89838_21	CAD	13.57
YGTCIYQGR	10	175	7	31.73	89838_16	CAD	21.43
LWAFCC	10	175	7	31.73	89838_16	CAD	20.09
IPACIAGER	10	175	8	33.65	89854_11	CAD	33.26

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
IPACIAGER	10	175	8	33.65	89854_15	CAD	25.36
IPACIAGER	10	175	8	33.65	89854_09	CAD	30.91
IPACIAGER	10	175	8	33.65	89854_13	CAD	27.05
YGTCIYQGR	10	175	8	33.65	89854_10	CAD	29.63
YGTCIYQGR	10	175	8	33.65	89854_09	CAD	30.73
YGTCIYQGR	10	175	9	35.7	89870_10	CAD	47.83
IPACIAGER	10	175	14	45.2	89866_08	CAD	27.9
IPACIAGER	11	175	7	31.73	88249_16	CAD	33.62
YGTCIYQGR	11	175	7	31.73	88249_17	CAD	26.13
YGTCIYQGR	11	175	7	31.73	88249_16	CAD	39.49
IPACIAGER	11	175	8	33.65	88265_09	CAD	39.79
IPACIAGER	11	175	8	33.65	88265_10	CAD	17.77
IPACIAGER	11	175	8	33.65	88265_13	CAD	21.59
IPACIAGER	11	175	8	33.65	88265_11	CAD	16.46
IPACIAGER	11	175	9	35.7	88277_10	CAD	29.52
YGTCIYQGR	11	175	9	35.7	88277_09	CAD	53.19
IPACIAGER	11	175	10	37.5	88205_07	CAD	25.87
IPACIAGER	11	175	13	43.27	88253_13	CAD	29.07
IPACIAGER	12	175	8	33.65	101029_07	CAD	16.26
IPACIAGER	12	175	8	33.65	101029_09	CAD	13.61
IPACIAGER	12	175	8	33.65	101029_06	CAD	10.92
YGTCIYQGR	12	175	8	33.65	101029_09	CAD	36.99
YGTCIYQGR	12	175	8	33.65	101029_07	CAD	31.12
YGTCIYQGR	12	175	8	33.65	101029_07	CAD	29.94
LWAFCC	12	175	8	33.65	101029_07	CAD	6.25
LWAFCC	12	175	8	33.65	101029_07	CAD	13.49
IPACIAGER	12	175	9	35.7	101053_09	CAD	20.24
IPACIAGER	12	175	9	35.7	101053_08	CAD	14.04
IPACIAGER	12	175	10	37.5	100965_08	CAD	26.08
YGTCIYQGR	12	175	10	37.5	100965_08	CAD	27.37
ADEVAAAPEQIAADIP EVVVSLAWDESLAPK	12	175	12	41.34	101001_09	CAD	69.47
ADEVAAAPEQIAADIP EVVVSLAWDESLAPK	12	175	14	45.2	101037_08	CAD	58.63

Table 1d – CPP 17							
Peptide Sequence	CEX	Salt	RP1	% B	Run	Proteome	Olav Score
IPACIAGER	13	175	8	33.65	89762_09	CAD	25.42
IPACIAGER	13	175	8	33.65	89762_14	CAD	28.91
YGTCIYQGR	13	175	8	33.65	89762_14	CAD	37
YGTCIYQGR	13	175	8	33.65	89762_13	CAD	37.54
LWAFCC	13	175	8	33.65	89762_09	CAD	14.67
IPACIAGER	13	175	9	35.7	89774_14	CAD	18.53
IPACIAGER	13	175	9	35.7	89774_13	CAD	40.87
YGTCIYQGR	13	175	9	35.7	89774_09	CAD	14.42
YGTCIYQGR	13	175	9	35.7	89774_10	CAD	15.03
LWAFCC	13	175	9	35.7	89774_09	CAD	18.93
LWAFCC	13	175	9	35.7	89774_10	CAD	17.74
YGTCIYQGR	13	175	10	37.5	89706_10	CAD	22.71
IPACIAGER	13	175	15	47.1	89786_09	CAD	24.05
IPACIAGER	14	175	9	35.7	92232_10	CAD	42.27
IPACIAGER	14	175	9	35.7	92232_11	CAD	25.42
LWAFCC	14	175	9	35.7	92232_10	CAD	21.28
IPACIAGER	14	175	10	37.5	92204_07	CAD	27.84
EVVVSLAWDESLAPK	14	175	12	41.34	92582_10	CAD	61.03
ADEVAAAPEQIAADIP EVVVSLAWDESLAPK	14	175	12	41.34	92582_09	CAD	70.47
IPACIAGER	14	175	13	43.27	92590_06	CAD	25.73
IPACIAGER	15	225	9	35.7	95617_12	CAD	17.02
IPACIAGER	15	225	9	35.7	95617_15	CAD	31.49
IPACIAGER	15	225	9	35.7	95617_14	CAD	29.94
IPACIAGER	15	225	9	35.7	95617_18	CAD	14.9
IPACIAGER	15	225	9	35.7	95617_17	CAD	21.35
IPACIAGER	15	225	9	35.7	95617_08	CAD	21.9
YGTCIYQGR	15	225	9	35.7	95617_16	CAD	23.54
YGTCIYQGR	15	225	9	35.7	95617_15	CAD	44.53
YGTCIYQGR	15	225	9	35.7	95617_10	CAD	38.13
LWAFCC	15	225	9	35.7	95617_09	CAD	15.3
LWAFCC	15	225	9	35.7	95617_09	CAD	18.93
YGTCIYQGR	18	1000	8	33.65	101101_08	CAD	51.06

For CPP 17, the ratio of protein levels in CAD versus control plasma samples is calculated by two methods. The first method calculates the CAD/Control ratio by the number of fractions from each sample containing CPP 17. The result is 2.4 (see Table 1d). Alternatively, and more accurately, the Olav scores obtained for each peptide in the mass spectrometry data analysis software are used to give a weighted ratio. This result is 2.1, indicating that the CPPs of the invention are present at 2.1 times the level in CAD plasma compared to control plasma. As such, the CPPs provide a useful diagnostic tool, wherein an increased level of a CPP indicates an increased risk of developing, or the presence of, a cardiovascular disorder.

Number of Control Fractions: 116
Total number of CAD Fractions: 279
Total CAD/ Control: 2.4

Olav Score Control: 3720.25
Olav Score Total CAD: 7946.59
Total CAD/ Control: 2.1

It is interesting to note that Table 1d displays the tryptic peptide of SEQ ID NO:18

NO:16, and which as such is not expected to be found in plasma. Besides, this tryptic peptide was identified in diseased plasma only. This observation may reflect an altered processing of the preprotein in the case of the disease.

Table 1e – CPP 20

CPP #	Peptide Sequence	CEX	Salt	RP1	% B	Run Number
CPP 20	AFQYHSK	5	75	9	35.7	101089_11
	CEEDKEFTCR	10	175	9	35.7	89870_11
	CEEDKEFTCR					89870_09
	EPLDDYVNTQGPSLFSVTK			10	37.5	89802_08
	CEEDKEFTCR					89802_08
	CEEDKEFTCR					89802_08
	EPLDDYVNTQGPSLFSVTK	11	175	10	37.5	88205_08
	EPLDDYVNTQGPSLFSVTK					88205_08
	CEEDKEFTCR					88205_08

EPLDDYVNTQGPSLFSVTK		11	39.42	88221_08

CPP nucleic acids

One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode CPPs or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic (DNA vaccine) and diagnostic methods and in drug screening assays as further described herein.

An object of the invention is a purified, isolated, or recombinant nucleic acid coding for a CPP, complementary sequences thereto, and fragments thereof. The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide coding for a CPP, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide coding for a CPP, or a sequence complementary thereto or a biologically active fragment thereof. Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide coding for a CPP, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

In another preferred aspect, the invention pertains to purified or isolated nucleic acid molecules that encode a portion or variant of a CPP, wherein the portion or variant displays a CPP biological activity. Preferably said portion or variant is a portion or variant of a naturally occurring CPP or precursor thereof.

Another object of the invention is a purified, isolated, or recombinant nucleic acid encoding a CPP comprising, consisting essentially of, or consisting of the amino acid sequence selected from one of the groups of SEQ ID Nos:1-5, 6-10, 11-14, 15-23 and 24-28, or fragments thereof, wherein the isolated nucleic acid molecule encodes one or more motifs such as, for CPP 2: a lipase binding region, a hydrophobic face, or a disulfide bond; for CPP 9: a substrate RNA-binding site, a substrate virion binding site (preferably a RSV virion), a ribonuclease active site, or a disulfide bond; for CPP 21: a cholesterol-binding domain, a glycosylation site, or a disulfide bond; for CPP 17: a bacterial endotoxin binding site or a disulfide bond; for CPP 20: a kringle-binding domain or a disulfide bond.

The nucleotide sequence determined from the cloning of the CPP-encoding gene allows for the generation of probes and primers designed for use in identifying and/or cloning other CPPs (e.g. sharing the novel functional domains), as well as CPP homologues from other species.

A nucleic acid fragment encoding a "biologically active portion of a CPP" can be prepared by

isolating a portion of a nucleotide sequence coding for a CPP, which encodes a polypeptide having a CPP biological activity, expressing the encoded portion of the CPP (e.g., by recombinant expression in vitro or in vivo) and assessing the activity of the encoded portion of the CPP.

The invention further encompasses nucleic acid molecules that differ from the CPP nucleotide sequences of the invention due to degeneracy of the genetic code and encode the same CPPs of the invention.

In addition to the CPP nucleotide sequences described above, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the CPPs may exist within a population (e.g., the human population). Such genetic polymorphism may exist among individuals within a population due to natural allelic variation. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a CPP-encoding gene or nucleic acid sequence.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the CPP nucleic acids of the invention can be isolated based on their homology to the CPP nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

It will be appreciated that the invention comprises polypeptides having an amino acid sequence encoded by any of the polynucleotides of the invention.

Uses of CPP nucleic acids

Polynucleotide sequences (or the complements thereof) encoding CPPs have various applications, including uses as hybridization probes, in chromosome and gene mapping, and in the generation of antisense RNA and DNA. In addition, CPP-encoding nucleic acids are useful as targets for pharmaceutical intervention, e.g. for the development of DNA vaccines, and for the preparation of CPPs by recombinant techniques, as described herein. The polynucleotides described herein, including sequence variants thereof, can be used in diagnostic assays. Accordingly, diagnostic methods based on detecting the presence of such polynucleotides in body fluids or tissue samples are a feature of the present invention. Examples of nucleic acid based diagnostic assays in accordance with the present invention include, but are not limited to, hybridization assays, e.g., in situ hybridization, and PCR-based assays. Polynucleotides, including extended length polynucleotides, sequence variants and fragments thereof, as described herein, may be used to generate hybridization probes or PCR primers for use in such assays. Such probes and primers will be capable of detecting polynucleotide sequences, including genomic sequences that are similar, or complementary to, the

CPP polynucleotides described herein.

The invention includes primer pairs for carrying out a PCR to amplify a segment of a polynucleotide of the invention. Each primer of a pair is an oligonucleotide having a length of between 15 and 30 nucleotides such that i) one primer of the pair forms a perfectly matched duplex with one strand of a polynucleotide of the invention and the other primer of the pair form a perfectly match duplex with the complementary strand of the same polynucleotide, and ii) the primers of a pair form such perfectly matched duplexes at sites on the polynucleotide that separated by a distance of between 10 and 2500 nucleotides. Preferably, the annealing temperature of each primer of a pair to its respective complementary sequence is substantially the same.

Hybridization probes derived from polynucleotides of the invention can be used, for example, in performing in situ hybridization on tissue samples, such as fixed or frozen tissue sections prepared on microscopic slides or suspended cells. Briefly, a labeled DNA or RNA probe is allowed to bind its DNA or RNA target sample in the tissue section on a prepared microscopic, under controlled conditions. Generally, dsDNA probes consisting of the DNA of interest cloned into a plasmid or bacteriophage DNA vector are used for this purpose, although ssDNA or ssRNA probes may also be used. Probes are generally oligonucleotides between about 15 and 40 nucleotides in length. Alternatively, the probes can be polynucleotide probes generated by PCR random priming primer extension or in vitro transcription of RNA from plasmids (riboprobes). These latter probes are typically several hundred base pairs in length. The probes can be labeled by any of a number of label groups and the particular detection method will correspond to the type of label utilized on the probe (e.g., autoradiography, X-ray detection, fluorescent or visual microscopic analysis, as appropriate). The reaction can be further amplified in situ using immunocytochemical techniques directed against the label of the detector molecule used, such as an antibody directed to a fluorescein moiety present on a fluorescently labeled probe. Specific labeling and in situ detection methods can be found, for example, in Howard, G. C., Ed., *Methods in Nonradioactive Detection,* Appleton & Lange, Norwalk, Conn., (1993), herein incorporated by reference.

Hybridization probes and PCR primers may also be selected from the genomic sequences corresponding to the full-length proteins identified in accordance with the present invention, including promoter, enhancer elements and introns of the gene encoding the naturally occurring polypeptide. Nucleotide sequences encoding a CPP can also be used to construct hybridization probes for mapping the gene encoding that CPP and for the genetic analysis of individuals. Individuals carrying variations of, or mutations in the gene encoding a CPP of the present invention may be detected at the DNA level by a variety of techniques. Nucleic acids used for diagnosis may be obtained from a patient's

cells, including, for example, tissue biopsy and autopsy material. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki, et al. *Nature* 324:163-166 (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid of the present invention can be used to identify and analyze mutations in the gene of the present invention. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA of the invention or alternatively, radiolabeled antisense DNA sequences of the invention. Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and Sl protection or the chemical cleavage method (e.g. Cotton, et al., *Proc. Natl. Acad. Sci. USA* 85:4397-4401 (1985)), or by differences in melting temperatures. "Molecular beacons" (Kostrikis L. G. et al., *Science* 279:1228-1229 (1998)), hairpin-shaped, single-stranded synthetic oligonucleotides containing probe sequences which are complementary to the nucleic acid of the present invention, may also be used to detect point mutations or other sequence changes as well as monitor expression levels of CPPs.

Oligonucleotide and Antisense Compounds

Oligonucleotides of the invention, including PCR primers and antisense compounds, are synthesized by conventional means or as commercially available automated DNA synthesizers (e.g. Applied Biosystems (Foster City, CA) model 380B, 392 or 394 DNA/RNA synthesizer, or like instrument. Preferably, phosphoramidite chemistry is employed, e.g. as disclosed in the following references: Beaucage and Iyer, Tetrahedron, 48: 2223-2311 (1992); Molko et al, U.S. patent 4,980,460; Koster et al, U.S. patent 4,725,677; Caruthers et al, U.S. patents 4,415,732; 4,458,066; and 4,973,679; and the like. For therapeutic use, nuclease resistant backbones are preferred. Many types of modified oligonucleotides are available that confer nuclease resistance, e.g. phosphorothioate, phosphorodithioate, phosphoramidate, or the like, described in many references, e.g. phosphorothioates: Stec et al, U.S. patent 5,151,510; Hirschbein, U.S. patent 5,166,387; Bergot, U.S. patent 5,183,885; phosphoramidates: Froehler et al, International application PCT/US90/03138; and for a review of additional applicable chemistries: Uhlmann and Peyman (cited above). The length of the antisense oligonucleotides has to be sufficiently large to ensure that specific binding will take place only at the desired target polynucleotide and not at other fortuitous sites. The upper range of the length is determined by several factors, including the inconvenience and expense of synthesizing and purifying oligomers greater than about 30-40 nucleotides in length, the greater tolerance of longer oligonucleotides for mismatches than shorter oligonucleotides, and the like. Preferably, the antisense

oligonucleotides of the invention have lengths in the range of about 15 to 40 nucleotides. More preferably, the oligonucleotide moieties have lengths in the range of about 18 to 25 nucleotides.

Primers and probes

Primers and probes of the invention can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang SA et al (Methods Enzymol 1979;68:90-98), the phosphodiester method of Brown EL et al (Methods Enzymol 1979;68:109-151), the diethylphosphoramidite method of Beaucage et al (Tetrahedron Lett 1981, 22: 1859-1862) and the solid support method described in EP 0 707 592, the disclosures of which are incorporated herein by reference in their entireties.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. If desired, the probe may be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating any label group known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. Additional examples include non-radioactive labeling of nucleic acid fragments as described in Urdea et al. (Nucleic Acids Research. 11:4937-4957, 1988) or Sanchez-Pescador et al. (J. Clin. Microbiol. 26(10):1934-1938, 1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al (Nucleic Acids Symp. Ser. 24:197-200, 1991) or in the European patent No. EP 0225807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore

depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in CPP-encoding genes or mRNA using other techniques.

Any of the nucleic acids, polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member attached to the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal,

glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes and other configurations known to those of ordinary skill in the art. The nucleic acids, polynucleotides, primers and probes of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on a solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips, and has been described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092, the disclosures of which are incorporated herein by reference in their entireties.

Methods for obtaining variant nucleic acids and polypeptides

In addition to naturally-occurring allelic variants of the CPP sequences that may exist in the population, the skilled artisan will appreciate that changes can be introduced by mutation into the nucleotide sequences coding for CPPs, thereby leading to changes in the amino acid sequence of the encoded CPPs, with or without altering the functional ability of the CPPs.

Several types of variants are contemplated including 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated CPP is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the CPP, such as a leader, a signal or anchor sequence, a sequence which is employed for purification of

the CPP, or sequence from a precursor protein. Such variants are deemed to be within the scope of those skilled in the art.

For example, nucleotide substitutions leading to amino acid substitutions can be made in the sequences that do not substantially change the biological activity of the protein. An amino acid residue can be altered from the wild-type sequence encoding a CPP, or a biologically active fragment or homologue thereof without altering the biological activity. In general, amino acid residues that are shared among the CPPs of the present invention are predicted to be less amenable to alteration.

In another aspect, the invention pertains to nucleic acid molecules encoding CPPs that contain changes in amino acid residues that result in increased biological activity, or a modified biological activity. In another aspect, the invention pertains to nucleic acid molecules encoding CPPs that contain changes in amino acid residues that are essential for a CPP biological activity. Such CPPs differ in amino acid sequence from SEQ ID NOs: 1-5, 6-10, 11-14, 15-23 and 24-28 and display reduced activity, or essentially lack one or more CPP biological activities.

Mutations, substitutions, additions, or deletions can be introduced into any of SEQ ID NOs: 1-5, 6-10, 11-14, 15-23 and 24-28, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. For example, conservative amino acid substitutions may be made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a CPP, or a biologically active fragment or homologue thereof may be replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a CPP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for CPP biological activity to identify mutants that retain activity. Following mutagenesis of the nucleotide encoding one of SEQ ID NOs: 1-5, 6-10, 11-14, 15-23 and 24-28, the encoded protein can be expressed recombinantly and the activity of the protein can be determined in any suitable assay, for example, as provided herein.

The invention also provides CPP chimeric or fusion proteins. As used herein, a CPP "chimeric

protein" or "fusion protein" comprises a CPP of the invention or fragment thereof, operatively linked or fused in frame to a non-CPP polypeptide sequence. In a preferred embodiment, a CPP fusion protein comprises at least one biologically active portion of a CPP. In another preferred embodiment, a CPP fusion protein comprises at least two biologically active portions of a CPP. For example, in one embodiment, the fusion protein is a GST-CPP fusion protein in which CPP domain sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant CPPs. In another embodiment, the fusion protein is a CPP containing a heterologous signal sequence at its N-terminus, for example, to allow for a desired cellular localization in a certain host cell. In yet another embodiment, the fusion is a CPP biologically active fragment and an immunoglobulin molecule. Such fusion proteins are useful, for example, to increase the valency of CPP binding sites. For example, a bivalent CPP binding site may be formed by fusing biologically active CPP fragments to an IgG Fc protein.

CPP fusion proteins of the invention can be used as immunogens to produce anti-CPP antibodies in a subject, to purify CPP or CPP ligands, and in screening assays to identify CPP modulators.

Furthermore, isolated fragments of CPPs can also be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a CPP of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments with a CPP biological activity, for example, by microinjection assays or in vitro protein binding assays. In an illustrative embodiment, peptidyl portions of a CPP, such as a CPP target binding region, can be tested for CPP activity by expression as thioredoxin fusion proteins, each of which contains a discrete fragment of the CPP (see, for example, U.S. Patents 5, 270,181 and 5,292,646; and PCT publication WO94/02502, the disclosures of which are incorporated herein by reference).

In addition, libraries of fragments of a CPP coding sequence can be used to generate a variegated population of CPP fragments for screening and subsequent selection of variants of a CPP. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of CPP coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products,

removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the CPP.

Whether a change in the amino acid sequence of a peptide results in a functional CPP homolog can be readily determined by assessing at least one CPP biological activity of the variant peptide. Peptides in which more than one replacement has taken place can readily be tested in the same manner.

Chemical Manufacture of CPP Compositions

Peptides of the invention are synthesized by standard techniques (e.g. Stewart and Young, Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Company, Rockford, IL, 1984). Preferably, a commercial peptide synthesizer is used, e.g. Applied Biosystems, Inc. (Foster City, CA) model 430A, and polypeptides of the invention may be assembled from multiple, separately synthesized and purified, peptide in a convergent synthesis approach, e.g. Kent et al, U.S. patent 6,184,344 and Dawson and Kent, Annu. Rev. Biochem., 69: 923-960 (2000). Peptides of the invention may be assembled by solid phase synthesis on a cross-linked polystyrene support starting from the carboxyl terminal residue and adding amino acids in a stepwise fashion until the entire peptide has been [illegible]. The following references are guides to the chemistry employed during synthesis: Solmajer et al, Int. J. Peptide Protein Res., 40: 180-193 (1992); Merrifield, J. Amer. Chem. Soc., Vol. 85, pg. 2149 (1963); Kent et al., pg 185, in Peptides 1984, Ragnarsson, Ed. (Almquist and Weksell, Stockholm, 1984); Kent et al., pg. 217 in Peptide Chemistry 84, Izumiya, Ed. (Protein Research Foundation, B.H. Osaka, 1985); Merrifield, Science, Vol. 232, pgs. 341-347 (1986); Kent, Ann. Rev. Biochem, Vol. 57, pgs. 957-989 (1988), and references cited in these latter two references.

Preferably, chemical synthesis of polypeptides of the invention is carried out by the assembly of peptide fragments by native chemical ligation, as described by Dawson et al, Science, 266: 776-779 (1994) and Kent el al, U.S. patent 6,184,344. Briefly, in the approach a first peptide fragment is provided with an N-terminal cysteine having an unoxidized sulfhydryl side chain, and a second peptide fragment is provided with a C-terminal thioester. The unoxidized sulfhydryl side chain of the N-terminal cysteine is then condensed with the C-terminal thioester to produce an intermediate peptide fragment which links the first and second peptide fragments with a β-aminothioester bond. The β-aminothioester bond of the intermediate peptide fragment then undergoes an intramolecular rearrangement to produce the peptide fragment product which links the first and second peptide

fragments with an amide bond. Preferably, the N-terminal cysteine of the internal fragments is protected from undesired cyclization and/or concatenation reactions by a cyclic thiazolidine protecting group as described below. Preferably, such cyclic thiazolidine protecting group is a thioprolinyl group.

Peptide fragments having a C-terminal thioester may be produced as described in the following references, which are incorporated by reference: Kent et al, U.S. patent 6,184,344; Tam et al, Proc. Natl. Acad. Sci., 92: 12485-12489 (1995); Blake, Int. J. Peptide Protein Res., 17: 273 (1981); Canne et al, Tetrahedron Letters, 36: 1217-1220 (1995); Hackeng et al, Proc. Natl. Acad. Sci., 94: 7845-7850 (1997); or Hackeng et al, Proc. Natl. Acad. Sci., 96: 10068-10073 (1999). Preferably, the method described by Hackeng et al (1999) is employed. Briefly, peptide fragments are synthesized on a solid phase support (described below) typically on a 0.25 mmol scale by using the in situ neutralization/HBTU activation procedure for Boc chemistry disclosed by Schnolzer et al, Int. J. Peptide Protein Res., 40: 180-193 (1992), which reference is incorporated herein by reference. (HBTU is 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate and Boc is tert-butoxycarbonyl). Each synthetic cycle consists of N^{α}-Boc removal by a 1- to 2- minute treatment with neat TFA, a 1-minute DMF flow wash, a 10- to 20-minute coupling time with 1.0 mmol of preactivated Boc-amino acid in the presence of DIEA, and a second DMF flow wash. (TFA is

Boc-amino acids (1.1 mmol) are preactivated for 3 minutes with 1.0 mmol of HBTU (0.5 M in DMF) in the presence of excess DIEA (3 mmol). After each coupling step, yields are determined by measuring residual free amine with a conventional quantitative ninhydrin assay, e.g. as disclosed in Sarin et al, Anal. Biochem., 117: 147-157 (1981). After coupling of Gln residues, a DCM flow wash is used before and after deprotection by using TFA, to prevent possible high-temperature (TFA/DMF)-catalyzed pyrrolidone formation. After chain assembly is completed, the peptide fragments are deprotected and cleaved from the resin by treatment with anhydrous HF for 1 hour at 0°C with 4% *p*-cresol as a scavenger. The imidazole side-chain 2,4-dinitrophenyl (dnp) protecting groups remain on the His residues because the dnp-removal procedure is incompatible with C-terminal thioester groups. However, dnp is gradually removed by thiols during the ligation reaction. After cleavage, peptide fragments are precipitated with ice-cold diethylether, dissolved in aqueous acetonitrile, and lyophilized.

Thioester peptide fragments described above are preferably synthesized on a trityl-associated mercaptopropionic acid-leucine (TAMPAL) resin, made as disclosed by Hackeng et al (1999), or comparable protocol. Briefly, N^{α}-Boc-Leu (4 mmol) is activated with 3.6 mmol of HBTU in the

presence of 6 mmol of DIEA and coupled for 16 minutes to 2 mmol of p-methylbenzhydrylamine (MBHA) resin, or the equivalent. Next, 3 mmol of S-trityl mercaptopropionic acid is activated with 2.7 mmol of HBTU in the presence of 6 mmol of DIEA and coupled for 16 minutes to Leu-MBHA resin. The resulting TAMPAL resin can be used as a starting resin for polypeptide-chain assembly after removal of the trityl protecting group with two 1-minute treatments with 3.5% triisopropylsilane and 2.5% H_2O in TFA. The thioester bond can be formed with any desired amino acid by using standard in situ-neutralization peptide coupling protocols for 1 hour, as disclosed in Schnolzer et al (cited above). Treatment of the final peptide fragment with anhydrous HF yields the C-terminal activated mercaptopropionic acid-leucine (MPAL) thioester peptide fragments.

Preferably, thiazolidine-protected thioester peptide fragment intermediates are used in native chemical ligation under conditions as described by Hackeng et al (1999), or like conditions. Briefly, 0.1 M phosphate buffer (pH 8.5) containing 6 M guanidine, 4% (vol/vol) benzylmercaptan, and 4% (vol/vol) thiophenol is added to dry peptides to be ligated, to give a final peptide concentration of 1-3 mM at about pH 7, lowered because of the addition of thiols and TFA from the lyophilized peptide. Preferably, the ligation reaction is performed in a heating block at 37°C and is periodically vortexed to equilibrate the thiol additives. The reaction may be monitored for degree of completion by MALDI-MS or HPLC and electrospray ionization MS.

ring of the product is opened by treatment with a cysteine deprotecting agent, such as O-methylhydroxylamine (0.5 M) at pH 3.5-4.5 for 2 hours at 37° C, after which a 10-fold excess of Tris-(2-carboxyethyl)-phosphine is added to the reaction mixture to completely reduce any oxidizing reaction constituents prior to purification of the product by conventional preparative HPLC. Preferably, fractions containing the ligation product are identified by electrospray MS, are pooled, and lyophilized.

After the synthesis is completed and the final product purified, the final polypeptide product may be refolded by conventional techniques, e.g. Creighton, Meth. Enzymol., 107: 305-329 (1984); White, Meth. Enzymol., 11: 481-484 (1967); Wetlaufer, Meth. Enzymol., 107: 301-304 (1984); and the like. Preferably, a final product is refolded by air oxidation by the following, or like: The reduced lyophilized product is dissolved (at about 0.1 mg/mL) in 1 M guanidine hydrochloride (or like chaotropic agent) with 100 mM Tris, 10 mM methionine, at pH 8.6. After gentle overnight stirring, the re-folded product is isolated by reverse phase HPLC with conventional protocols.

Recombinant Expression Vectors and Host Cells

The polynucleotide sequences described herein can be used in recombinant DNA molecules that direct the expression of the corresponding polypeptides in appropriate host cells. Because of the degeneracy in the genetic code, other DNA sequences may encode the equivalent amino acid sequence, and may be used to clone and express the CPPs. Codons preferred by a particular host cell may be selected and substituted into the naturally occurring nucleotide sequences, to increase the rate and/or efficiency of expression. The nucleic acid (e.g., cDNA or genomic DNA) encoding the desired CPP may be inserted into a replicable vector for cloning (amplification of the DNA), or for expression. The polypeptide can be expressed recombinantly in any of a number of expression systems according to methods known in the art (Ausubel, et al., editors, Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1990). Appropriate host cells include yeast, bacteria, archebacteria, fungi, and insect and animal cells, including mammalian cells, for example primary cells, including stem cells, including, but not limited to bone marrow stem cells. More specifically, these include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors, and yeast transformed with yeast expression vectors. Also included, are insect cells infected with a recombinant insect virus (such as baculovirus), and mammalian expression systems. The nucleic acid sequence to be expressed may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an

generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The CPPs of the present invention are produced by culturing a host cell transformed with an expression vector containing a nucleic acid encoding a CPP, under the appropriate conditions to induce or cause expression of the protein. The conditions appropriate for CPP expression will vary with the choice of the expression vector and the host cell, as ascertained by one skilled in the art. For example, the use of constitutive promoters in the expression vector may require routine optimization of host cell growth and proliferation, while the use of an inducible promoter requires the appropriate growth conditions for induction. In addition, in some embodiments, the timing of the harvest is important. For example, the baculoviral systems used in insect cell expression are lytic viruses, and thus harvest time selection can be crucial for product yield.

A host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein

include, but are not limited to, glycosyl, acetyl, phosphate, amide, lipid, carboxyl, acyl, or carbohydrate groups. Post-translational processing, which cleaves a "prepro" form of the protein, may also be important for correct insertion, folding and/or function. By way of example, host cells such as CHO, HeLa, BHK, MDCK, 293, W138, etc. have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein. Of particular interest are *Drosophila melanogaster* cells, *Saccharomyces cerevisiae* and other yeasts, *E. coli, Bacillus subtilis,* SF9 cells, Cl29 cells, 293 cells, Neurospora, BHK, CHO, COS, and HeLa cells, fibroblasts, Schwanoma cell lines, immortalized mammalian myeloid and lymphoid cell lines, Jurkat cells, human cells and other primary cells.

The nucleic acid encoding a CPP must be "operably linked" by placing it into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" DNA sequences are contiguous, and, in the case of a secretory leader or other polypeptide sequence, contiguous and in

at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice. Promoter sequences encode either constitutive or inducible promoters. The promoters may be either naturally occurring promoters or hybrid promoters. Hybrid promoters, which combine elements of more than one promoter, are also known in the art, and are useful in the present invention. The expression vector may comprise additional elements, for example, the expression vector may have two replication systems, thus allowing it to be maintained in two organisms, for example in mammalian or insect cells for expression and in a procaryotic host for cloning and amplification. Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2: plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells. Further, for integrating expression vectors, the expression vector contains at least one sequence homologous to the host cell genome, and preferably, two homologous sequences which flank the expression construct. The integrating vector may be

directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors are well known in the art. In an additional embodiment, a heterologous expression control element may be operably linked with the endogenous gene in the host cell by homologous recombination (described in US Patents 6410266 and 6361972, disclosures of which are hereby incorporated by reference in their entireties). This technique allows one to regulate expression to a desired level with a chosen control element while ensuring proper processing and modification of CPP endogenously expressed by the host cell. Useful heterologous expression control elements include but are not limited to CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous Sarcoma Virus (RSV), and metallothionein promoters.

Preferably, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selection genes are well known in the art and will vary with the host cell used. Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available for from complex media, e.g., the gene encoding D-alanine racemase for Bacilli.

conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding the CPP can be designed with signal sequences which direct secretion of the CPP through a prokaryotic or eukaryotic cell membrane. The desired CPP may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the CPP-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, e.g., the yeast invertase leader, alpha factor leader (including Saccharomyces and Kluyveromyces a-factor leaders, the latter described in U.S. Pat. No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published Apr. 4, 1990), or the signal described in WO 90113646 published Nov. 15, 1990. In

mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders. According to the expression system selected, the coding sequence is inserted into an appropriate vector, which in turn may require the presence of certain characteristic "control elements" or "regulatory sequences." Appropriate constructs are known generally in the art (Ausubel, et al., 1990) and, in many cases, are available from commercial suppliers such as Invitrogen (San Diego, Calif.), Stratagene (La Jolla, Calif.), Gibco BRL (Rockville, Md.) or Clontech (Palo Alto, Calif.).

Expression in Bacterial Systems

Transformation of bacterial cells may be achieved using an inducible promoter such as the hybrid lacZ promoter of the "BLUESCRIPT" Phagemid (Stratagene) or "pSPORT1" (Gibco BRL). In addition, a number of expression vectors may be selected for use in bacterial cells to produce cleavable fusion proteins that can be easily detected and/or purified, including, but not limited to "BLUESCRIPT" (a-galactosidase; Stratagene) or pGEX (glutathione S-transferase; Promega, Madison, Wis.). A suitable bacterial promoter is any nucleic acid sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of the coding sequence of

placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose and maltose, and sequences derived from biosynthetic enzymes such as tryptophan. Promoters from bacteriophage may also be used and are known in the art. In addition, synthetic promoters and hybrid promoters are also useful; for example, the tat promoter is a hybrid of the trp and lac promoter sequences. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. An efficient ribosome-binding site is also desirable. The expression vector may also include a signal peptide sequence that provides for secretion of the CPP in bacteria. The signal sequence typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell, as is well known in the art. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). The bacterial expression vector may also include a selectable marker

gene to allow for the selection of bacterial strains that have been transformed. Suitable selection genes include drug resistance genes such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin and tetracycline. Selectable markers also include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways. When large quantities of CPPs are needed, e.g., for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the CPP coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; PIN vectors (Van Heeke & Schuster *JBiol Chem 264:5503-5509* 1989)); PET vectors (Novagen, Madison Wis.); and the like. Expression vectors for bacteria include the various components set forth above, and are well known in the art. Examples include vectors for *Bacillus subtilis, E. coli, Streptococcus cremoris,* and *Streptococcus lividans,* among others. Bacterial expression vectors are transformed into bacterial host cells using techniques well known in the art, such as calcium chloride mediated transfection, electroporation, and others.

Expression in Yeast

Saccharomyces cerevisiae, Candida albicans and *C. maltosa, Hansenula polymorpha, Kluyveromyces fragilis* and *K. lactis, Pichia guillermondii* and *P pastoris, Schizosaccharomyces pombe,* and *Yarrowia lipolytica.* Examples of suitable promoters for use in yeast hosts include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem.* 255:2073 (1980)) or other glycolytic enzymes (Hess et al., J. *Adv. Enzyme Reg.* 7:149 (1968); Holland, *Biochemistry* 17:4900 (1978)), such as enolase, glyceraldehyde-3- phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose- 6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, alpha factor, the ADH2IGAPDH promoter, glucokinase alcohol oxidase, and PGH. See, for example, Ausubel, et al., 1990; Grant et al., *Methods in Enzymology* 153:516-544, (1987). Other yeast promoters, which are inducible have the additional advantage of transcription controlled by growth conditions, include the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657. Yeast selectable markers include ADE2. HIS4. LEU2. TRPl. and

ALG7, which confers resistance to tunicamycin; the neomycin phosphotransferase gene, which confers resistance to G418; and the CUP1 gene, which allows yeast to grow in the presence of copper ions. Yeast expression vectors can be constructed for intracellular production or secretion of a CPP from the DNA encoding the CPP of interest. For example, a selected signal peptide and the appropriate constitutive or inducible promoter may be inserted into suitable restriction sites in the selected plasmid for direct intracellular expression of the CPP. For secretion of the CPP, DNA encoding the CPP can be cloned into the selected plasmid, together with DNA encoding the promoter, the yeast alpha-factor secretory signal/leader sequence, and linker sequences (as needed), for expression of the CPP. Yeast cells, can then be transformed with the expression plasmids described above, and cultured in an appropriate fermentation media. The protein produced by such transformed yeast can then be concentrated by precipitation with 10% trichloroacetic acid and analyzed following separation by SDS-PAGE and staining of the gels with Coomassie Blue stain. The recombinant CPP can subsequently be isolated and purified from the fermentation medium by techniques known to those of skill in the art.

Expression in Mammalian Systems

The CPP may be expressed in mammalian cells. Mammalian expression systems are known in

transformed with any of a number of different viral-based expression systems, such as adenovirus, where the coding region can be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential El or E3 region of the viral genome results in a viable virus capable of expression of the polypeptide of interest in infected host cells. A preferred expression vector system is a retroviral vector system such as is generally described in PCT/US97/01019 and PCT/US97/101048. Suitable mammalian expression vectors contain a mammalian promoter which is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence for CPP into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, using a located 25-30 base pairs upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element (enhancer element), typically located within 100 to 200 base pairs upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation. Of particular use as mammalian promoters are the promoters from mammalian viral

genes, since the viral genes are often highly expressed and have a broad host range. Examples include promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211, 504 published Jul. 5,1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems. Transcription of DNA encoding a CPP by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, a-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer is preferably located at a site 5' from the promoter. In general, the transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-translational cleavage and polyadenylation. Examples of transcription terminator and

recombinant proteins can be effected in a stable expression system. Expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene may be used for this purpose. Appropriate vectors containing selectable markers for use in mammalian cells are readily available commercially and are known to persons skilled in the art. Examples of such selectable markers include, but are not limited to herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase for use in tk- or hprt-cells, respectively. The methods of introducing exogenous nucleic acid into mammalian hosts, as well as other hosts, is well known in the art, and will vary with the host cell used. Techniques include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, viral infection, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

CPPs can be purified from culture supernatants of mammalian cells transiently transfected or stably transformed by an expression vector carrying a CPP-encoding sequence. Preferably, CPP is purified from culture supernatants of COS 7 cells transiently transfected by the pcD expression vector. Transfection of COS 7 cells with pcD proceeds as follows: One day prior to transfection,

approximately 10^6 COS 7 monkey cells are seeded onto individual 100 mm plates in Dulbecco's modified Eagle medium (DME) containing 10% fetal calf serum and 2 mM glutamine. To perform the transfection, the medium is aspirated from each plate and replaced with 4 ml of DME containing 50 mM Tris.HCl pH 7.4, 400 mg/ml DEAE-Dextran and 50 μg of plasmid DNA. The plates are incubated for four hours at 37°C, then the DNA-containing medium is removed, and the plates are washed twice with 5 ml of serum-free DME. DME is added back to the plates which are then incubated for an additional 3 hrs at 37°C. The plates are washed once with DME, after which DME containing 4% fetal calf serum, 2 mM glutamine, penicillin (100 U/L) and streptomycin (100 μg/L) at standard concentrations is added. The cells are then incubated for 72 hrs at 37°C, after which the growth medium is collected for purification of CPP. Plasmid DNA for the transfections is obtained by growing pcD(SRα), or like expression vector, containing the CPP-encoding cDNA insert in E. coli MC1061 (described by Casadaban and Cohen, J. Mol. Biol., Vol. 138, pgs. 179-207 (1980)), or like organism. The plasmid DNA is isolated from the cultures by standard techniques, e.g. Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor Laboratory, New York, 1989) or Ausubel et al (1990, cited above).

CPPs may also be produced in insect cells. Expression vectors for the transformation of insect cells, and in particular, baculovirus-based expression vectors, are well known in the art. In one such system, the CPP-encoding DNA is fused upstream of an epitope tag contained within a baculovirus expression vector. *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* Sf9 cells or in Trichoplusia larvae. The CPP-encoding sequence is cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of a CPP-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect *S. frugiperda* cells or Trichoplusia larvae in which the CPP is expressed (Smith et al., J. Wol. 46:584 (1994); Engelhard E K et al., *Proc. Nat. Acad. Sci.* 91:3224-3227 (1994)). Suitable epitope tags for fusion to the CPP-encoding DNA include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including commercially available plasmids such as pVL1393 (Novagen). Briefly, the CPP-encoding DNA or the desired portion of the CPP-encoding DNA is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking restriction sites. The PCR product is then digested with the selected restriction enzymes and subcloned into an expression

vector. Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGoldTM virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL), or other methods known to those of skill in the art. Virus is produced by day 4-5 of culture in Sf9 cells at 28°C, and used for further amplifications. Procedures are performed as further described in O'Reilley et al., *BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL*, Oxford University Press (1994). Extracts may be prepared from recombinant virus-infected Sf9 cells as described in Rupert et al., *Nature* 362:175-179 (1993). Alternatively, expressed epitope-tagged CPP can be purified by affinity chromatography, or for example, purification of an IgG tagged (or Fc tagged) CPP can be performed using chromatography techniques, including Protein A or protein G column chromatography.

Evaluation of Gene Expression

Gene expression may be evaluated in a sample directly, for example, by standard techniques known to those of skill in the art, e.g., Northern blotting to determine the transcription of mRNA, dot blotting (DNA or RNA), or in situ hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be used in assays for detection of polypeptides, nucleic acids, such as specific duplexes, including DNA duplexes, RNA duplexes, and

carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected. Gene expression, alternatively, may be measured by immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to directly evaluate the expression of a CPP polypeptide or polynucleotide. Antibodies useful for such immunological assays may be either monoclonal or polyclonal, and may be prepared against a native sequence CPP. Protein levels may also be detected by mass spectrometry. A further method of protein detection is with protein chips.

Purification of Expressed Protein

Expressed CPP may be purified or isolated after expression, using any of a variety of methods known to those skilled in the art. The appropriate technique will vary depending upon what other components are present in the sample. Contaminant components that are removed by isolation or purification are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other solutes. The purification step(s) selected will depend, for example, on the nature of the production process used and the particular CPP

produced. As CPPs are secreted, they may be recovered from culture medium. Alternatively, the CPP may be recovered from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (e.g. Triton-X 100) or by enzymatic cleavage. Alternatively, cells employed in expression of CPP can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or by use of cell lysing agents. Exemplary purification methods include, but are not limited to, ion-exchange column chromatography; chromatography using silica gel or a cation-exchange resin such as DEAE; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; chromatography using metal chelating columns to bind epitope-tagged forms of the CPP; ethanol precipitation; reverse phase HPLC; chromatofocusing; SDS-PAGE; and ammonium sulfate precipitation. Ordinarily, an isolated CPP will be prepared by at least one purification step. For example, the CPP may be purified using a standard anti-CPP antibody column. Ultrafiltration and dialysis techniques, in conjunction with protein concentration, are also useful (see, for example, Scopes, R., *PROTEIN PURIFICATION,* Springer-Verlag, New York, N.Y., 1982). The degree of purification necessary will vary depending on the use of the CPP. In some instances no purification will be necessary. Once expressed and purified as needed, the CPPs and nucleic acids of the present invention are useful in a number of applications, as detailed herein.

Transgenic animals

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which CPP-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous CPP sequences have been introduced into their genome or homologous recombinant animals in which endogenous CPP sequences have been altered. Such animals are useful for studying the function and/or activity of a CPP or fragment thereof and for identifying and/or evaluating modulators of CPP biological activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal include a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal,

more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a CPP-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection or retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The CPP cDNA sequence or a fragment thereof can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human CPP-encoding gene, such as from mouse or rat, can be used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a CPP transgene to direct expression of a CPP to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Pat. No. 4,873,191 by Wagner et al. and in Hogan, B., Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986, the disclosure of which is incorporated herein by reference in its entirety). Similar methods are used for production of other transgenic animals. A transgenic

expression of CPP mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a CPP can further be bred to other transgenic animals carrying other transgenes.

To create an animal in which a desired nucleic acid has been introduced into the genome via homologous recombination, a vector is prepared which contains at least a portion of a CPP-encoding sequence into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the CPP-encoding sequence. The CPP-encoding sequence can be a human gene, but more preferably, is a non-human homologue of a human CPP-encoding sequence (e.g., a cDNA isolated by stringent hybridization with a nucleotide sequence coding for a CPP). For example, a mouse CPP-encoding sequence can be used to construct a homologous recombination vector suitable for altering an endogenous gene in the mouse genome. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous CPP-encoding sequence is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous CPP-encoding sequence is mutated or otherwise altered but still encodes functional

protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous CPP-encoding sequence). In the homologous recombination vector, the altered portion of the CPP-encoding sequence is flanked at its 5' and 3' ends by additional nucleic acid sequence of the CPP gene to allow for homologous recombination to occur between the exogenous sequence carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K. R. and Capecchi, M. R. (1987) Cell 51:503, the disclosure of which is incorporated herein by reference in its entirety, for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced CPP-encoding sequence has homologously recombined with the endogenous gene are selected (see e.g., Li, E. et al. (1992) Cell 69:915, the disclosure of which is incorporated herein by reference in its entirety). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in Teratocarcinomas and Embryonic Stem Cells. A Practical Approach, E. J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152, the disclosure of which is incorporated herein by reference in its entirety). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to

animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) Current Opinion in Biotechnology 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec et al.; WO 91/01140 by Smithies et al.; WO 92/0968 by Zijlstra et al.; and WO 93/04169 by Berns et al., the disclosures of which are incorporated herein by reference in their entireties.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, e.g., Lakso et al. (1992) PNAS 89:6232-6236, the disclosure of which is incorporated herein by reference in its entirety. Another example of a recombinase system is the FLP recombinase system of Saccharomyces cerevisiae (O'Gorman et al. (1991) Science 251:1351-1355, the disclosure of which is incorporated herein by reference in its entirety). If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the

construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Assessing CPP activity

It will be appreciated that the invention further provides methods of testing the activity of or obtaining functional fragments and variants of CPPs and CPP sequences. Such methods involve providing a variant or modified CPP-encoding nucleic acid and assessing whether the encoded polypeptide displays a CPP biological activity. Encompassed is thus a method of assessing the function of a CPP comprising: (a) providing a CPP, or a biologically active fragment or homologue thereof; and (b) testing said CPP, or a biologically active fragment or homologue thereof for a CPP biological activity under conditions suitable for CPP activity. Cell free, cell-based and in vivo assays may be used to test CPP activity. For example, said assay may comprise expressing a CPP nucleic acid in a host cell, and observing CPP activity in said cell and other affected cells. In another example, a CPP, or a biologically active fragment or homologue thereof is contacted with a cell, and a CPP biological activity is observed.

CPP biological activities include: (1) indicating that an individual has or will have a cardiovascular disorder; (2) circulating through the bloodstream of individuals with a cardiovascular

the ability to generate an anti-CPP specific antibody; and for CPP 2: (5) interacting with a CPP target protein, preferably a lipase; (6) stabilizing the active site of a lipase; (7) increasing lipase activity; (8) interacting with a CPP target molecule such as a phospholipid, micelle, or triglyceride; and (9) forming at least one disulfide bond; for CPP 9: (5) forming intramolecular amino acid side chain interactions such as hydrogen, amide, or especially disulfide links; (6) interaction with a CPP target molecule, preferably an RNA molecule or virion (such as respiratory syncytial virus or RSV); (7) antiviral activity, and (8) hydrolysis of RNA phosphodiester bonds; for CPP 21: (5) forming intramolecular amino acid side chain interactions such as hydrogen, amide, or preferably disulfide links; and (6) interacting with a CPP target molecule, in particular, cholesterol; for CPP 17: (5) forming intramolecular amino acid side chain interactions such as hydrogen, amide, or preferably disulfide links; (6) interaction with a CPP target molecule, preferably a low-density lipoprotein or bacterial endotoxin; (7) neutralizing bacterial endotoxins; (8) promoting mast cell chemotaxis; (9) undergoing posttranslational processing, for example, specific proteolysis; (10) functioning as an antimicrobial defense; and (11) inhibiting contraction of smooth muscle cells; and for CPP 20: (5) forming intramolecular amino acid side chain interactions such as hydrogen, amide, or preferably

disulfide links; (6) interacting with a CPP target molecule, in particular, a kringle domain-containing peptides such as plasminogen; and (7) reducing tumor growth.

CPP biological activity can be assayed by any suitable method known in the art. Antigenicity and immunogenicity may be detected, for example, as described in the sections titled "Anti CPP antibodies" and "Uses of CPP antibodies". Circulation in blood plasma may be detected as described in "Diagnostic and Prognostic Uses". Interaction with a CPP target molecule may be detected according to any of the methods described herein, for example, in the section titled "Drug Screening Assays".

Determining the ability of the CPP to bind to or interact with a CPP target molecule can be accomplished by a method for directly or indirectly determining binding, as is common to the art. Such methods can be cell-based (e.g., such that binding to a membrane-bound CPP is detected) or cell free. Interaction of a test compound with a CPP can be detected, for example, by coupling the CPP or biologically active portion thereof with a label group such that binding of the CPP or biologically active portion thereof to its cognate target molecule can be determined by detecting the labeled CPP or biologically active portion thereof in a complex. For example, the extent of complex formation may be measured by immunoprecipitating the complex or by performing gel electrophoresis. Determining the ability of the CPP to bind to a CPP target molecule may also be accomplished using a

C. (1991) Anal. Chem. 63:2338-2345 and Szabo et al. (1995) Curr. Opin. Struct. Biol. 5:699-705, the disclosures of which are incorporated herein by reference in their entireties. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

For CPP 2, Lipase activity may be determined by any assay capable of detecting the hydrolysis reaction catalyzed by lipase. For example, a reduction in the amount of substrate or, alternatively, an increase in the amount of product, in a test sample can be measured. A preferred assay is described by Panteghini, et al. (Ann Clin Biochem (2001) 38:365-370, disclosures of which are incorporated herein in its entirety). Briefly, a chromogenic substrate, DGGR, is provided in an assay mixture with lipase, bile salts, calcium, and colipase. Lipase activity on the substrate is measured by detecting the formation of methylresorufin, the chromophore product, at 580nm. Lipase activity is ideally measured to detect the CPP biological activities of stabilizing the active site of a lipase and increasing lipase activity. An example of such a method comprises the steps of: contacting a biological sample with an assay mixture of chromogenic substrate and lipase under conditions

suitable for lipase activity; and detecting lipase activity. Preferably, the biological sample is plasma, more preferably plasma from an individual suspected of having a cardiovascular disorder. CPP activity is indicated by an increase in lipase activity when compared to an appropriate negative control.

For CPP 9, RNase activity may be assayed by the following exemplary techniques. Enzyme is incubated 15 min at 37C in Buffer B (1mM reduced glutathione (Sigma Chemical Co., St Louis, MO), 0.5 mM oxidized glutathione, 0.1 mM DTT, 0.5 mM Tris-HCl 8.2, 0.01 mM EDTA, and 0.1M urea) with 1ug yeast tRNA (Sigma) as a substrate. Electrophoresis is carried out in a 1.5% agarose gel in 0.09 M Tris-phosphate, 0.002 M EDTA, and 0.5 microg/ml ethidium bromide. RNA is visualized by UV. RNase activity leads to the disappearance of the full-length RNA. Alternatively, RNase activity may be detected spectrophotometrically. Briefly, the concentration of perchloric acid-soluble ribonucleotides generated from acid-insoluble yeast tRNA (Sigma) in a 40 mM sodium phosphate, pH 7.5, buffer by a given quantity of RNase is measured spectrophotometrically at 260 nm. For example, an RNase-containing sample corresponding to approximately 100 nM recombinant RNase is added in a 0.8 ml reaction volume with 10 µl of 4 mg/ml yeast tRNA. Increasing RNase activity results in reduced 260 nm values.

For CPP 9, antiviral activity may be detected as follows. Briefly, CPP 9-containing sample at 10^3 infectious units/ml) in culture medium (Iscove's Modified Dulbecco's Medium with 10% heat-inactivated fetal calf serum and 2 mM glutamine) and incubated with gentle rotation at room temperature. After 2 h incubation, 200 microl of the suspension is used to infect target cells (human respiratory epithelial HEp-2 cells, in the case of respiratory syncytial virus) present in confluent monolayers (3-4 × 10^5 cells/monolayer) on coverslips within a one dram shell vial (Viromed, Minneapolis, MN). After spin amplification (700 *g* at 22°C) and 16 h incubation (37°C, 5% CO_2), the primary infected cells are identified by immunofluorescent staining (mouse anti-RSV blend, FITC-labeled; Chemicon International, Temencula, CA). Data is presented as infectious units ± SD.

For CPP 17, interaction with an LDL particle may be detected according to the methods of Higazi, et al. Blood 96:1393-1398 (2000), relevant disclosure of which is incorporated herein. Briefly, the peptides of interest were labelled with radioactive iodine. LDL or a BSA control (10nM) was incubated with 0-5 uM ^{125}I-defensin for 1hr at 37C. The mixture was applied to a polyacrylamide gel and the amount of radioactivity in the appropriate molecular weight fraction recorded. Specific LDL binding was defined by the increased amount of radioactive peptide bound to LDL compared to BSA. Antimicrobial activity may be detected according to any method known in the art, such as those

described by Porter et al. (Infect.Immun. 65: 2396-2401 (1997)). Briefly, *L. monocytogenes* EGD, *E. coli* ML35p, *S. typhimurium* 14028S and 7953S, and *C. albicans* 820 are used. The bacterial cultures are harvested in midgrowth phase, washed, and resuspended at a working dilution of 10^6 bacteria/ml in 10 mM sodium phosphate (pH7.4), 1% Trypticase Soy Broth. *C. albicans* is harvested in the same manner but isolated in stationary phase. Colony Forming Unit (CFU) assays and Radial diffusion assays are performed as described in the presence of a test compound, a positive control compound, or no compound. A reduction in the value obtained for either the CFU or radial diffusion assays in the presence of a test compound as compared to the control without compound indicates that the compound has antimicrobial activity.

For CPP 2, CPP 9, CPP 17, CPP 20 and CPP 21, intramolecular interactions may be detected by sequence-based structural predictions. Such predictions are generally based on X-ray crystallography or NMR structural data for a polypeptide with similar sequence. Detection of intramolecular interactions may also be accomplished using SDS-PAGE. For the example of disulfide bonds, links formed between different portions of a given protein result in a more compacted protein, and thus, a reduced apparent molecular weight. Disulfide bonds may be disrupted by a reducing agent, for example, dithiothreitol (DTT). A protein sample that has been treated with a reducing agent may thus be compared to an untreated control by SDS-PAGE to detect a change in apparent

For CPP 17, antimicrobial activity may be detected according to any method known in the art, such as those described by Porter et al. (Infect.Immun. 65: 2396-2401 (1997)). Briefly, *L. monocytogenes* EGD, *E. coli* ML35p, *S. typhimurium* 14028S and 7953S, and *C. albicans* 820 are used. The bacterial cultures are harvested in midgrowth phase, washed, and resuspended at a working dilution of 10^6 bacteria/ml in 10 mM sodium phosphate (pH7.4), 1% Trypticase Soy Broth. *C. albicans* is harvested in the same manner but isolated in stationary phase. Colony Forming Unit (CFU) assays and Radial diffusion assays are performed as described in the presence of a test compound, a positive control compound, or no compound. A reduction in the value obtained for either the CFU or radial diffusion assays in the presence of a test compound as compared to the control without compound indicates that the compound has antimicrobial activity.

For CPP 17, specific proteolysis may be detected by comparing the molecular weight of a sample peptide to that of a peptide of known molecular weight. Molecular weights are easily compared according to any method common to the art such as SDS-PAGE, gel chromatography, or mass spectrometry. Preferably, the molecular weight of a test peptide is obtained by mass spectrometry and compared to a database comprising molecular weights of peptides with

posttranslational modifications. Exemplary databases include Genpept, SWISSPROT, EMBL, and the Protein Sequence Database. Such techniques are detailed further herein.

For CPP 17, chemotactic activity may be assessed as described for T cells in U.S. Patent 5837247. Briefly, lymphocyte migration was assessed using a 48 well microchemotaxis chamber (Neuro Probe Inc. Cabin John, Md.). 25 ul of the sample to be tested diluted in chemotaxis medium was placed in the lower compartment and 50 ul of cell suspension (at 5×10^6 cells/ ml) in the upper compartment. The two compartments were separated by a polycarbonate filter (for T cells, 5 μm pore size, larger for other lymphocytes) coated with 10 ug/ml collagen type IV overnight at 4C. The apparatus was incubated at 37C for 3 h in humidified air with 5% CO_2. At the end of the incubation period, the filter was removed, fixed and stained with LeukoStat (Fisher Scientific, Pittsburgh, Pa.). The number of cells that migrate through the filter were counted by light microscopy per high-power fields. The results are expressed as the mean (±SD) value of the migration in triplicate samples and are representative of at least three experiments. The statistical significance of the number of cells migrating in response to stimuli versus control medium was calculated using the Student's T test. For the in vivo assay, BALB/C and CB-17 scid/scid (SCID) mice were obtained from the Animal Production Area (NCI-FCRDC, Frederick, Md.). Mice were used at 8-12 wk of age and kept in pathogen free conditions. SCID mice were treated with anti-ASGM-1 and injected with 1×10^8 huPBL

PBS were injected daily subcutaneously into the same injection site. The injection site was examined histologically either at 4 hr after the first or at 24 hr by 4 hr after a second injection. Experiments were performed on three to four mice per group in duplicate.

For CPP 17, vascular smooth muscle contraction may be measured as described in Nassar, et al. (Blood 100:4026-32 (2002)) or EP0582631B1, relevant disclosures of which are incorporated by reference. Generally, aortic ring tissue sections from rat or another suitable animal are used for the assay. The tissue section rings are treated with, for example, epinephrine, a powerful contractile stimulant and the lengths are measured in the presence and absence of a test substance. An increased length of the section relative to the negative control indicates that a substance reduces contractile activity.

For CPP 20, any method known in the art may be used to assay for tumor growth. For example, in the experiments detailed in WO99/46282 (relevant disclosures of which are incorporated herein by reference), C57B16/J mice were implanted with Lewis lung carcinomas. A suspension of 10^6 tumor cells in 100 microliters of PBS was injected into the subcutaneous dorsa of the mice. Tumors were measured with a dial-caliper, and volumes were determined using a general formula for

the volume of an ellipsoid sphere (width^2 x length x 0.52). The mice were randomized into two groups once the tumor volume reached about 160 mm3. One group was treated with the test substance and the other received a PBS control. Tumor growth was monitored over 11 days.

Cardiovascular disorders may be diagnosed by any method determined appropriate for an individual by one of skill in the art. Further examples of symptoms and diagnostics may be found in the Background section, and are best determined appropriately by one of skill in the art based on the particular profile of a patient.

Intramolecular interactions may be detected by sequence-based structural predictions. Such predictions are generally based on X-ray crystallography or NMR structural data for a polypeptide with similar sequence. Detection of intramolecular interactions may also be accomplished using SDS-PAGE. For the example of disulfide bonds, links formed between different portions of a given protein result in a more compacted protein, and thus, a reduced apparent molecular weight. Disulfide bonds may be disrupted by a reducing agent, for example, dithiothreitol (DTT). A protein sample that has been treated with a reducing agent may thus be compared to an untreated control by SDS-PAGE to detect a change in apparent molecular weight. Such methods are common to the art.

The present invention provides antibodies and binding compositions specific for CPPs. Such antibodies and binding compositions include polyclonal antibodies, monoclonal antibodies, Fab and single chain Fv fragments thereof, bispecific antibodies, heteroconjugates, and humanized antibodies. Such antibodies and binding compositions may be produced in a variety of ways, including hybridoma cultures, recombinant expression in bacteria or mammalian cell cultures, and recombinant expression in transgenic animals. There is abundant guidance in the literature for selecting a particular production methodology, e.g. Chadd and Chamow, Curr. Opin. Biotechnol., 12: 188-194 (2001).

The choice of manufacturing methodology depends on several factors including the antibody structure desired, the importance of carbohydrate moieties on the antibodies, ease of culturing and purification, and cost. Many different antibody structures may be generated using standard expression technology, including full-length antibodies, antibody fragments, such as Fab and Fv fragments, as well as chimeric antibodies comprising components from different species. Antibody fragments of small size, such as Fab and Fv fragments, having no effector functions and limited pharmokinetic activity may be generated in a bacterial expression system. Single chain Fv fragments are highly selective for in vivo tumors, show good tumor penetration and low immunogenicity, and

are cleared rapidly from the blood, e.g. Freyre et al, J. Biotechnol., 76: 157-163 (2000). Thus, such molecules are desirable for radioimmunodetection.

Polyclonal Antibodies

The anti-CPP antibodies of the present invention may be polyclonal antibodies. Such polyclonal antibodies can be produced in a mammal, for example, following one or more injections of an immunizing agent, and preferably, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected into the mammal by a series of subcutaneous or intraperitoneal injections. The immunizing agent may include CPPs or a fusion protein thereof. It may be useful to conjugate the antigen to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include, but are not limited to, keyhole limpet hemocyanin (KLH), methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Adjuvants include, for example, Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicoryno-mycolate). The immunization protocol may be determined by one skilled in the art based on standard protocols or by routine experimentation.

Alternatively, a crude protein preparation which has been enriched for a CPP or a portion

into the non-human mammal in the presence of an appropriate adjuvant. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies are purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate and excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appear to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987), the disclosure of which is incorporated herein by reference in its entirety. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. J. Clin. Endocrinol. Metab. 33:988-991(1971), the disclosure of which is incorporated by reference in its entirety. Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973). Plateau concentration of

antibody is usually in the range of 0.1 to 0.2 mg/ml of serum. Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D. C. (1980).

Monoclonal Antibodies

Alternatively, the anti-CPP antibodies may be monoclonal antibodies. Monoclonal antibodies may be produced by hybridomas, wherein a mouse, hamster, or other appropriate host animal, is immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent, e.g. Kohler and Milstein, Nature 256:495 (1975). The immunizing agent will typically include the CPP or a fusion protein thereof and optionally a carrier. Alternatively, the lymphocytes may be immunized in vitro. Generally, spleen cells or lymph node cells are used if non-human mammalian sources are desired, or peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired. The lymphocytes are fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to produce a hybridoma cell, e.g. Goding, *MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE,* Academic Press, pp. 59-103 (1986); Liddell and Cryer, A Practical Guide to Monoclonal Antibodies

Press, New York, 1994). In general, immortalized cell lines are transformed mammalian cells, for example, myeloma cells of rat, mouse, bovine or human origin. The hybridoma cells are cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine (HAT), substances which prevent the growth of HGPRT-deficient cells. Preferred immortalized cell lines are those that fuse efficiently, support stable high level production of antibody, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine or human myeloma lines, which can be obtained, for example, from the American Type Culture Collection (ATCC), Rockville, MD. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies, e.g. Kozbor, J. Immunol. 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, pp. 51-63 (1987).

The culture medium (supernatant) in which the hybridoma cells are cultured can be assayed for the presence of monoclonal antibodies directed against a CPP. Preferably, the binding specificity

of monoclonal antibodies present in the hybridoma supernatant is determined by immunoprecipitation or by an in vitro binding assay, such as radio- immunoassay (RIA) or Enzyme-Linked Immuno Sorbent Assay (ELISA). Appropriate techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.* 107:220 (1980). After the desired antibody-producing hybridoma cells are identified, the cells may be cloned by limiting dilution procedures and grown by standard methods (Goding, 1986, *supra*). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown in vivo as ascites in a mammal. The monoclonal antibodies secreted by selected clones may be isolated or purified from the culture medium or ascites fluid by immunoglobulin purification procedures routinely used by those of skill in the art such as, for example, protein A-Sepharose, hydroxyl-apatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Pat. No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be isolated from the CPP-specific hybridoma cells and sequenced, e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies. Once isolated, the DNA may be inserted into an expression vector, which is then

cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for the murine heavy and light chain constant domains for the homologous human sequences (Morrison et al., *Proc. Nat. Acad. Sci.* 81:6851-6855 (1984); Neuberger et al., *Nature* 312:604-608 (1984); Takeda et al., *Nature* 314:452-454 (1985)), or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. The non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody. The antibodies may also be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, in vitro methods are suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

Antibodies and antibody fragments characteristic of hybridomas of the invention can also be produced by recombinant means by extracting messenger RNA, constructing a cDNA library, and

selecting clones which encode segments of the antibody molecule. The following are exemplary references disclosing recombinant techniques for producing antibodies: Wall et al., Nucleic Acids Research, Vol. 5, pgs. 3113-3128 (1978); Zakut et al., Nucleic Acids Research, Vol. 8, pgs. 3591-3601 (1980); Cabilly et al., Proc. Natl. Acad. Sci., Vol. 81, pgs. 3273-3277 (1984); Boss et al., Nucleic Acids Research, Vol. 12, pgs. 3791-3806 (1984); Amster et al., Nucleic Acids Research, Vol. 8, pgs. 2055-2065 (1980); Moore et al., U.S. Patent 4,642,334; Skerra et al, Science, Vol. 240, pgs. 1038-1041(1988); Huse et al, Science, Vol. 246, pgs. 1275-1281 (1989); and U.S. patents 6,054,297; 5,530,101; 4,816,567; 5,750,105; and 5,648,237; which patents are incorporated by reference. In particular, such techniques can be used to produce interspecific monoclonal antibodies, wherein the binding region of one species is combined with non-binding region of the antibody of another species to reduce immunogenicity, e.g. Liu et al., Proc. Natl. Acad. Sci., Vol. 84, pgs. 3439-3443 (1987), and patents 6,054,297 and 5,530,101. Preferably, recombinantly produced Fab and Fv fragments are expressed in bacterial host systems. Preferably, full-length antibodies are produced by mammalian cell culture techniques. More preferably, full-length antibodies are expressed in Chinese Hamster Ovary (CHO) cells or NSO cells.

Both polyclonal and monoclonal antibodies can be screened by ELISA. As in other solid phase immunoassays, the test is based on the tendency of macromolecules to adsorb nonspecifically to formation of antigen-antibody complexes with a simple separation of such complexes from unbound material. To titrate anti-peptide serum, peptide conjugated to a carrier different from that used in immunization is adsorbed to the wells of a 96-well microtiter plate. The adsorbed antigen is then allowed to react in the wells with dilutions of anti-peptide serum. Unbound antibody is washed away, and the remaining antigen-antibody complexes are allowed to react with an antibody specific for the IgG of the immunized animal. This second antibody is conjugated to an enzyme such as alkaline phosphatase. A visible colored reaction produced when the enzyme substrate is added indicates which wells have bound antipeptide antibodies. The use of spectrophotometer readings allows better quantification of the amount of peptide-specific antibody bound. High-titer antisera yield a linear titration curve between 10^{-3} and 10^{-5} dilutions.

CPP peptide carriers

The invention includes immunogens derived from CPPs, and immunogens comprising conjugates between carriers and peptides of the invention. The term immunogen as used herein refers to a substance which is capable of causing an immune response. The term carrier as used herein

refers to any substance which when chemically conjugated to a peptide of the invention permits a host organism immunized with the resulting conjugate to generate antibodies specific for the conjugated peptide. Carriers include red blood cells, bacteriophages, proteins, or synthetic particles such as agarose beads. Preferably, carriers are proteins, such as serum albumin, gamma-globulin, keyhole limpet hemocyanin (KLH), thyroglobulin, ovalbumin, fibrinogen, or the like.

The general technique of linking synthetic peptides to a carrier is described in several references, e.g. Walter and Doolittle, "Antibodies Against Synthetic Peptides," in Setlow et al., eds., Genetic Engineering, Vol. 5, pgs. 61-91 (Plenum Press, N.Y., 1983); Green et al. Cell, Vol. 28, pgs. 477-487 (1982); Lerner et al., Proc. Natl. Acad. Sci., Vol. 78, pgs. 3403-3407 (1981); Shimizu et al., U.S. Patent 4,474,754; and Ganfield et al., U.S. Patent 4,311,639. Accordingly, these references are incorporated by reference. Also, techniques employed to link haptens to carriers are essentially the same as the above-referenced techniques, e.g. chapter 20 in Tijssen, Practice and Theory of Enzyme Immunoassays (Elsevier, New York, 1985). The four most commonly used schemes for attaching a peptide to a carrier are (1) glutaraldehyde for amino coupling, e.g. as disclosed by Kagan and Glick, in Jaffe and Behrman, eds. Methods of Hormone Radioimmunoassay, pgs. 328-329 (Academic Press, N.Y., 1979), and Walter et al. Proc. Natl. Acad. Sci., Vol. 77, pgs. 5197-5200 (1980); (2) water-soluble carbodiimides for carboxyl to amino coupling, e.g. as disclosed by Hoare et al., J. Biol.

sidechain coupling, e.g. as disclosed by Bassiri et al., pgs. 46-47, in Jaffe and Behrman, eds. (cited above), and Walter et al. (cited above); and (4) maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) for coupling cysteine (or other sulfhydryls) to amino groups, e.g. as disclosed by Kitagawa et al., J. Biochem. (Tokyo), Vol. 79, pgs. 233-239 (1976), and Lerner et al. (cited above). A general rule for selecting an appropriate method for coupling a given peptide to a protein carrier can be stated as follows: the group involved in attachment should occur only once in the sequence, preferably at the appropriate end of the segment. For example, BDB should not be used if a tyrosine residue occurs in the main part of a sequence chosen for its potentially antigenic character. Similarly, centrally located lysines rule out the glutaraldehyde method, and the occurrences of aspartic and glutamic acids frequently exclude the carbodiimide approach. On the other hand, suitable residues can be positioned at either end of chosen sequence segment as attachment sites, whether or not they occur in the "native" protein sequence. Internal segments, unlike the amino and carboxy termini, will differ significantly at the "unattached end" from the same sequence as it is found in the native protein where the polypeptide backbone is continuous. The problem can be remedied, to a degree, by acetylating the α-amino group and then attaching the peptide by way of its carboxy terminus. The coupling

efficiency to the carrier protein is conveniently measured by using a radioactively labeled peptide, prepared either by using a radioactive amino acid for one step of the synthesis or by labeling the completed peptide by the iodination of a tyrosine residue. The presence of tyrosine in the peptide also allows one to set up a sensitive radioimmune assay, if desirable. Therefore, tyrosine can be introduced as a terminal residue if it is not part of the peptide sequence defined by the native polypeptide.

Preferred carriers are proteins, and preferred protein carriers include bovine serum albumin, myoglobulin, ovalbumin (OVA), keyhole limpet hemocyanin (KLH), or the like. Peptides can be linked to KLH through cysteines by MBS as disclosed by Liu et al., Biochemistry, Vol. 18, pgs. 690-697 (1979). The peptides are dissolved in phosphate-buffered saline (pH 7.5), 0.1 M sodium borate buffer (pH 9.0) or 1.0 M sodium acetate buffer (pH 4.0). The pH for the dissolution of the peptide is chosen to optimize peptide solubility. The content of free cysteine for soluble peptides is determined by Ellman's method, Ellman, Arch. Biochem. Biophys., Vol. 82, pg. 7077 (1959). For each peptide, 4 mg KLH in 0.25 ml of 10 mM sodium phosphate buffer (pH 7.2) is reacted with 0.7 mg MBS (dissolved in dimethyl formamide) and stirred for 30 min at room temperature. The MBS is added dropwise to ensure that the local concentration of formamide is not too high, as KLH is insoluble in >30% formamide. The reaction product, KLH-MBS, is then passed through Sephadex G-25

peak fractions of the column eluate (monitored by OD280) is estimated to be approximately 80%. KLH-MBS is then reacted with 5 mg peptide dissolved in 1 ml of the chosen buffer. The pH is adjusted to 7-7.5 and the reaction is stirred for 3 hr at room temperature. Coupling efficiency is monitored with radioactive peptide by dialysis of a sample of the conjugate against phosphate-buffered saline, and may range from 8% to 60%. Once the peptide-carrier conjugate is available, polyclonal or monoclonal antibodies are produced by standard techniques, e.g. as disclosed by Campbell, Monoclonal Antibody Technology (Elsevier, New York, 1984); Hurrell, ed. Monoclonal Hybridoma Antibodies: Techniques and Applications (CRC Press, Boca Raton, FL, 1982); Schreier et al. Hybridoma Techniques (Cold Spring Harbor Laboratory, New York, 1980); U.S. Patent 4,562,003; or the like. In particular, U.S. Patent 4,562,003 is incorporated by reference.

Humanized Antibodies

The anti-CPP antibodies of the invention may further comprise humanized antibodies or human antibodies. The term "humanized antibody" refers to humanized forms of non-human (e.g., murine) antibodies that are chimeric antibodies, immunoglobulin chains or fragments thereof (such as

Fv, Fab, Fab', F(ab'), or other antigen-binding partial sequences of antibodies) which contain some portion of the sequence derived from non-human antibody. Humanized antibodies include human immunoglobulins in which residues from a complementary determining region (CDR) of the human immunoglobulin are replaced by residues from a CDR of a non-human species such as mouse, rat or rabbit having the desired binding specificity, affinity and capacity. In general, the humanized antibody will comprise substantially all of at least one, and generally two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., *Nature* 321:522-525 (1986) and Presta, *Curr. Op. Struct. Biol. 2:593-596* (1992)). Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acids introduced into it from a source which is non-human in order to more closely resemble a human antibody, while still retaining the original binding activity of the antibody. Methods for humanization of antibodies are further detailed in Jones et al., *Nature* 321:522-525 (1986); Riechmann et al., *Nature* 332:323-327 (1988); and Verhoeyen et al., *Science* 239:1534-1536 (1988). Such "humanized" antibodies are chimeric antibodies in that substantially less than an intact human variable domain has been

Heteroconjugate Antibodies

Heteroconjugate antibodies which comprise two covalently joined antibodies, are also within the scope of the present invention. Heteroconjugate antibodies may be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be prepared using a disulfide exchange reaction or by forming a thioether bond.

Bispecific Antibodies

Bispecific antibodies have binding specificities for at least two different antigens. Such antibodies are monoclonal, and preferably human or humanized. One of the binding specificities of a bispecific antibody of the present invention is for a CPP, and the other one is preferably for a cell-surface protein or receptor or receptor subunit. Methods for making bispecific antibodies are known in the art, and in general, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs in hybridoma cells, where the two heavy chains have different specificities, e.g. Milstein and Cuello, *Nature 305:537-539* (1983). Given

that the random assortment of immunoglobulin heavy and light chains results in production of potentially ten different antibody molecules by the hybridomas, purification of the correct molecule usually requires some sort of affinity purification, e.g. affinity chromatography.

Uses of CPP antibodies

CPP antibodies are preferably specific for the CPPs of the invention and, as such, do not bind peptides derived from other proteins with high affinity. As used herein, the term "heavy chain variable region" means a polypeptide (1) which is from 110 to 125 amino acids in length, and (2) whose amino acid sequence corresponds to that of a heavy chain of an antibody of the invention, starting from the heavy chain's N-terminal amino acid. Likewise, the term "light chain variable region" means a polypeptide (1) which is from 95 to 115 amino acids in length, and (2) whose amino acid sequence corresponds to that of a light chain of an antibody of the invention, starting from the light chain's N-terminal amino acid. As used herein the term "monoclonal antibody" refers to homogeneous populations of immunoglobulins which are capable of specifically binding to CPPs.

CPP antibodies may be used as functional modulators, preferably as antagonists. Preferably, antibody modulators of the invention are derived from monoclonal antibodies specific for CPPs. Monoclonal antibodies capable of blocking, or neutralizing, CPPs are selected by their ability to

The use of antibody fragments is also well known, e.g. Fab fragments: Tijssen, Practice and Theory of Enzyme Immunoassays (Elsevier, Amsterdam, 1985); and Fv fragments: Hochman et al. Biochemistry, Vol. 12, pgs. 1130-1135 (1973), Sharon et al., Biochemistry, Vol. 15, pgs. 1591-1594 (1976) and Ehrlich et al., U.S. Patent 4,355,023; and antibody half molecules: Auditore- Hargreaves, U.S. Patent 4,470,925.

Preferably, monoclonal antibodies, Fv fragments, Fab fragments, or other binding compositions derived from monoclonal antibodies of the invention have a high affinity to CPPs. The affinity of monoclonal antibodies and related molecules to CPPs may be measured by conventional techniques including plasmon resonance, ELISA, or equilibrium dialysis. Affinity measurement by plasmon resonance techniques may be carried out, for example, using a BIAcore 2000 instrument (Biacore AB, Uppsala, Sweden) in accordance with the manufacturer's recommended protocol. Preferably, affinity is measured by ELISA, as described in U.S. patent 6,235,883, for example. Preferably, the dissociation constant between CPPs and monoclonal antibodies of the invention is less than 10^{-5} molar. More preferably, such dissociation constant is less than 10^{-8} molar; still more preferably, such dissociation constant is less than 10^{-9} molar; and most preferably, such dissociation

constant is in the range of 10^{-9} to 10^{-11} molar.

In addition, the antibodies of the present invention are useful for detecting CPPs. Such detection methods are advantageously applied to diagnosis of cardiovascular disorders, in particular, coronary artery disease. The antibodies of the invention may be used in most assays involving antigen-antibody reactions. The assays may be homogeneous or heterogeneous. In a homogeneous assay approach, the sample can be a biological sample or fluid such as serum, urine, whole blood, lymphatic fluid, plasma, saliva, cells, tissue, and material secreted by cells or tissues cultured in vitro. The sample can be pretreated if necessary to remove unwanted materials. The immunological reaction usually involves the specific antibody, a labeled analyte, and the sample suspected of containing the antigen. The signal arising from the label is modified, directly or indirectly, upon the binding of the antibody to the labeled analyte. Both immunological reaction and detection of the extent thereof are carried out in a homogeneous solution. Immunochemical labels which may be employed include free radicals, fluorescent dyes, enzymes, bacteriophages, coenzymes, and so forth.

In a heterogeneous assay approach, the reagents are usually the sample, the specific antibody, and means for producing a detectable signal. The specimen is generally placed on a support, such as a plate or a slide, and contacted with the antibody in a liquid phase. The support is then separated from the liquid phase and either the support phase or the liquid phase is examined for a detectable signal

presence of the antigen in the sample. Means for producing a detectable signal includes the use of radioactive labels, fluorescent compounds, enzymes, and so forth. Exemplary heterogeneous immunoassays are the radioimmunoassay, immunofluorescence methods, enzyme-linked immunoassays, and the like.

For a more detailed discussion of the above immunoassay techniques, see "Enzyme-Immunoassay," by Edward T. Maggio, CRC Press, Inc., Boca Raton, Fla., 1980. See also, for example, U.S. Pat. Nos. 3,690,834; 3,791,932; 3,817,837; 3,850,578; 3,853,987; 3,867,517; 3,901,654; 3,935,074; 3,984,533; 3,966,345; and 4,098,876, which listing is not intended to be exhaustive. Methods for conjugating labels to antibodies and antibody fragments are well known in the art. Such methods may be found in U.S. Pat. Nos. 4,220,450; 4,235,869; 3,935,974; and 3,966,345. Another example of a technique in which the antibodies of the invention may be employed is immunoperoxidase labeling. (Sternberger, Immunocytochemistry (1979) pp. 104-169). Alternatively, the antibodies may be bound to a radioactive material or to a drug to form a radiopharmaceutical or pharmaceutical, respectively. (Carrasquillo, et al., Cancer Treatment Reports (1984) 68:317-328).

One embodiment of an assay employing an antibody of the present invention involves the use of a surface to which the monoclonal antibody of the invention is attached. The underlying structure of the surface may take different forms, have different compositions and may be a mixture of compositions or laminates or combinations thereof. The surface may assume a variety of shapes and forms and may have varied dimensions, depending on the manner of use and measurement. Illustrative surfaces may be pads, beads, discs, or strips which may be flat, concave or convex. Thickness is not critical, generally being from about 0.1 to 2 mm thick and of any convenient diameter or other dimensions. The surface typically will be supported on a rod, tube, capillary, fiber, strip, disc, plate, cuvette and will typically be porous and polyfunctional or capable of being polyfunctionalized so as to permit covalent binding of an antibody and permit bonding of other compounds which form a part of a means for producing a detectable signal. A wide variety of organic and inorganic polymers, both natural and synthetic, and combinations thereof, may be employed as the material for the solid surface. Illustrative polymers include polyethylene, polypropylene, poly(4-methylbutene), polystyrene, polymethracrylate, poly(ethylene terephthalate), rayon, nylon, poly(vinyl butyrate), silicones, polyformaldehyde, cellulose, cellulose acetate, nitrocellulose, and latex. Other surfaces include paper, glasses, ceramics, metals, metaloids, semiconductor materials, cements, silicates, or the like. Also included are substrates that form gels, gelatins, lipopolysaccharides, silicates, agarose and

glycols (alkylene of 2 to 3 carbon atoms) or surfactants such as phospholipids. The binding of the antibody to the surface may be accomplished by well known techniques, commonly available in the literature (see, for example, "Immobilized Enzymes," Ichiro Chibata, Press, New York (1978) and Cuatrecasas, J. Bio. Chem., 245: 3059 (1970)). In carrying out the assay in accordance with this aspect of the invention, the sample is mixed with aqueous medium and the medium is contacted with the surface having an antibody bound thereto. Labels may be included in the aqueous medium, either concurrently or added subsequently so as to provide a detectable signal associated with the surface. The means for producing the detectable signal can involve the incorporation of a labeled analyte or it may involve the use of a second monoclonal antibody having a label conjugated thereto. Separation and washing steps will be carried out as needed. The signal detected is related to the presence of CPP in the sample. It is within the scope of the present invention to include a calibration on the same support. A particular embodiment of an assay in accordance with the present invention, by way of illustration and not limitation, involves the use of a support such as a slide or a well of a petri dish. The technique involves fixing the sample to be analyzed on the support with an appropriate fixing material and incubating the sample on the slide with a monoclonal antibody. After washing with an

appropriate buffer such as, for example, phosphate buffered saline, the support is contacted with a labeled specific binding partner for the antibody. After incubation as desired, the slide is washed a second time with an aqueous buffer and the determination is made of the binding of the labeled monoclonal antibody to the antigen. If the label is fluorescent, the slide may be covered with a fluorescent antibody mounting fluid on a cover slip and then examined with a fluorescent microscope to determine the extent of binding. On the other hand, the label can be an enzyme conjugated to the monoclonal antibody and the extent of binding can be determined by examining the slide for the presence of enzyme activity, which may be indicated by the formation of a precipitate, color, etc. A particular example of an assay utilizing the present antibodies is a double determinant ELISA assay. A support such as, e.g., a glass or vinyl plate, is coated with an antibody specific for CPP by conventional techniques. The support is contacted with the sample suspected of containing CPP, usually in aqueous medium. After an incubation period from 30 seconds to 12 hours, the support is separated from the medium, washed to remove unbound CPP with, for example, water or an aqueous buffered medium, and contacted with an antibody specific for CPP, again usually in aqueous medium. The antibody is labeled with an enzyme directly or indirectly such as, e.g., horseradish peroxidase or alkaline phosphatase. After incubation, the support is separated from the medium, and washed as above. The enzyme activity of the support or the aqueous medium is determined. This enzyme activity

For an example of immunoassays detecting the CPP 9 polypeptides of the invention, see Kurokawa, E. et al., Clin.Chim.Acta (1983), 128(1):83-93.

The invention also includes kits, e.g., diagnostic assay kits, for carrying out the methods disclosed above. In one embodiment, the kit comprises in packaged combination (a) a monoclonal antibody more specifically defined above and (b) a conjugate of a specific binding partner for the above monoclonal antibody and a label capable of producing a detectable signal. The reagents may also include ancillary agents such as buffering agents and protein stabilizing agents, e.g., polysaccharides and the like. The kit may further include, where necessary, other members of the signal producing system of which system the label is a member, agents for reducing background interference in a test, control reagents, apparatus for conducting a test, and the like. In another embodiment, the diagnostic kit comprises a conjugate of monoclonal antibody of the invention and a label capable of producing a detectable signal. Ancillary agents as mentioned above may also be present.

Further, an anti-CPP antibody (e.g., monoclonal antibody) can be used to isolate CPPs by standard techniques, such as affinity chromatography or immunoprecipitation. For example, an anti-

CPP antibody can facilitate the purification of natural CPPs from cells and of recombinantly produced CPP expressed in host cells. Moreover, an anti-CPP antibody can be used to isolate CPP to aid in detection of low concentrations of CPP (e.g., in plasma, cellular lysate or cell supernatant) or in order to evaluate the abundance and pattern of expression of the CPP. Anti-CPP antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a label group.

Protein Arrays

Detection, purification, and screening of the polypeptides of the invention may be accomplished using retentate chromatography (preferably, protein arrays or chips), as described by U.S. Patent 6225027 and U.S. Patent Application 20010014461, disclosures of which are herein incorporated by reference in their entireties. Briefly, retentate chromatography describes methods in which polypeptides (and/ or other sample components) are retained on an adsorbent (e.g., array or chip) and subsequently detected. Such methods involve (1) selectively adsorbing polypeptides from a sample to a substrate under a plurality of different adsorbent/eluant combinations ("selectivity conditions") and (2) detecting the retention of adsorbed polypeptides by desorption spectrometry (e.g.,

adsorbent prior to detection. The coupling of adsorption chromatography with detection by desorption spectrometry provides extraordinary sensitivity, the ability to rapidly analyze retained components with a variety of different selectivity conditions, and parallel processing of components adsorbed to different sites (i.e., "affinity sites" or "spots") on the array under different elution conditions.

These methods are useful for: combinatorial, biochemical separation and purification of the CPPs; study of differential gene expression; detection of differences in protein levels (e.g., for diagnosis); and detection of molecular recognition events, (e.g., for screening and drug discovery). Thus, this invention provides a molecular discovery and diagnostic device that is characterized by the inclusion of both parallel and multiplex polypeptide processing capabilities. Polypeptides of the invention and CPP-binding substances are preferably attached to a label group, and thus directly detected, enabling simultaneous transmission of two or more signals from the same "circuit" (i.e., addressable "chip" location) during a single unit operation.

Detection of CPPs by mass spectrometry

In accordance with the present invention, any instrument, method, process, etc. can be utilized to determine the identity and abundance of proteins in a sample. A preferred method of obtaining identity is by mass spectrometry, where protein molecules in a sample are ionized and then the resultant mass and charge of the protein ions are detected and determined.

To use mass spectrometry to analyze proteins, it is preferred that the protein be converted to a gas-ion phase. Various methods of protein ionization are useful, including, e.g., fast ion bombardment (FAB), plasma desorption, laser desorption, thermal desorption, preferably, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). Many different mass analyzers are available for peptide and protein analysis, including, but not limited to, Time-of-Flight (TOF), ion trap (ITMS), Fourier transform ion cyclotron (FTMS), quadrupole ion trap, and sector (electric and/or magnetic) spectrometers. See, e.g., U.S. Pat. No. 5,572,025 for an ion trap MS. Mass analyzers can be used alone, or in combination with other mass analyzers in tandem mass spectrometers. In the latter case, a first mass analyzer can be use to separate the protein ions (precursor ion) from each other and determine the molecular weights of the various protein constituents in the sample. A second mass analyzer can be used to analyze each separated constituents, e.g., by fragmenting the precursor ions into product ions by using, e.g. an inert gas. Any desired combination of mass analyzers can be used, including, e.g., triple quadrupoles, tandem time-of-flights, ion traps, and/or combinations thereof.

detectors can be utilized, such as ion electron multipliers or cryogenic detectors (e.g., U.S. Pat. No. 5,640,010). Additionally, non-destructive detectors can be used, such as ion traps which are used as ion current pick-up devices in quadrupole ion trap mass analyzers or FTMS.

For MALDI-TOF, a number of sample preparation methods can be utilized including, dried droplet (Karasand Hillenkamp, Anal. Chem., 60:2299-2301, 1988), vacuum-drying (Winberger et al., In Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, May 31-June 4, 1993, pp. 775a-b), crush crystals (Xiang et al., Rapid Comm. Mass Spectrom., 8:199-204,1994), slow crystal growing (Xiang et al., Org. Mass Spectrom, 28:1424-1429, 1993); active film (Mock et al., Rapid Comm. Mass Spectrom.,6:233-238, 1992; Bai et al., Anal. Chem., 66:3423-3430, 1994), pneumatic spray (Kochling et al., Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics; Atlanta, GA, May 21-26, 1995, p1225); electrospray (Hensel et al., Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics; Atlanta, GA, May 21 -26, 1995, p947); fast solvent evaporation (Vorm et al., Anal. Chem., 66:3281-3287, 1994); sandwich (Li et al., J. Am. Chem. Soc., 11 8:11662-11663,1996); and two-layer methods (Dal et al., Anal. Chem., 71:1087-1091, 1999). See also, e.g., Liang et al.,

Rapid Commun. Mass Spectrom., 10: 1219-1226, 1996; van Adrichemet al., Anal. Chem., 70:923-930, 1998.

For MALDI analysis, samples are prepared as solid-state co-crystals or thin films by mixing them with an energy absorbing compound or colloid (the matrix) in the liquid phase, and ultimately drying the solution to the solid state upon the surface of an inert probe. In some cases an energy absorbing molecule (EAM) is an integral component of the sample presenting surface. Regardless of EAM application strategy, the probe contents are allowed to dry to the solid state prior to introduction into the laser desorption/ionization time-of-flight mass spectrometer (LDIMS).

Ion detection in TOF mass spectrometry is typically achieved with the use of electro-emissive detectors such as electron multipliers (EMP) or microchannel plates (MCP). Both of these devices function by converting primary incident charged particles into a cascade of secondary, tertiary, quaternary, etc. electrons. The probability of secondary electrons being generated by the impact of a single incident charged particle can be taken to be the ion-to-electron conversion efficiency of this charged particle (or more simply, the conversion efficiency). The total electron yield for cascading events when compared to the total number of incident charged particles is typically described as the detector gain. Because generally the overall response time of MCPs is far superior to that of EMPs, MCPs are the preferred electro-emissive detector for enhancing mass/charge resolving power.

response time and broad frequency bandwidth are not necessary.

In a preferred aspect, for the analysis of digested proteins, a liquid-chromatography tandem mass spectrometer (LC-TMS) is used. This system provides an additional stage of sample separation via use of a liquid chromatograph followed by tandem mass spectrometry.

In preferred aspects, a protein eluted from a column according to the system described in Example 1 is analyzed using both MS and MS-MS analysis. For example, a small portion of intact proteins eluting from RP2 may be diverted to online detection using LC-ESI MS. The proteins are aliquoted on a number of plates allowing digestion or not with trypsin, preparation for MALDI-MS as well as for ESI-MS, as well as preparation of the MALDI plates with different matrices. The methods thus allow, in addition to information on intact mass, to conduct an analysis by both peptide mass fingerprinting and MS-MS techniques.

The methods described herein of separating and fractionating proteins provide individual proteins or fractions containing small numbers of distinct proteins. These proteins can be identified by mass spectral determination of the molecular masses of the protein and peptides resulting from the fragmentation thereof. Making use of available information in protein sequence databases, a

comparison can be made between proteolytic peptide mass patterns generated *in silico*, and experimentally observed peptide masses. A "hit-list" can be compiled, ranking candidate proteins in the database, based on (among other criteria) the number of matches between the theoretical and experimental proteolytic fragments. Several Web sites are accessible that provide software for protein identification on-line, based on peptide mapping and sequence database search strategies (e.g., http://www.expasy.ch). Methods of peptide mapping and sequencing using MS are described in WO 95/252819, U.S. Pat. No. 5,538,897, U.S. Pat. No. 5,869,240, U.S. Pat. No. 5,572,259, and U.S. Pat. No. 5,696,376. See, also, Yates, J. Mass Spec., 33:1 (1998).

Data collected from a mass spectrometer typically comprises the intensity and mass to charge ratio for each detected event. Spectral data can be recorded in any suitable form, including, e.g., in graphical, numerical, or electronic formats, either in digital or analog form. Spectra are preferably recorded in a storage medium, including, e.g., magnetic, such as floppy disk, tape, or hard disk; optical, such as CD-ROM or laser-disc; or, ROM-CHIPS.

The mass spectrum of a given sample typically provides information on protein intensity, mass to charge ratio, and molecular weight. In preferred embodiments of the invention, the molecular weights of proteins in the sample are used as a matching criterion to query a database. The molecular weights are calculated conventionally, e.g., by subtracting the mass of the ionizing proton for singly-

charges for multiply-charged ions and subtracting the number of ionizing protons.

Various databases are useful in accordance with the present invention. Useful databases include, databases containing genomic sequences, expressed gene sequences, and/or expressed protein sequences. Preferred databases contain nucleotide sequence-derived molecular masses of proteins present in a known organism, organ, tissue, or cell-type. There are a number of algorithms to identify open reading frames (ORF) and convert nucleotide sequences into protein sequence and molecular weight information. Several publicly accessible databases are available, including, the SwissPROT/TrEMBL database (http://www.expasy.ch).

Typically, a mass spectrometer is equipped with commercial software that identifies peaks above a certain threshold level, calculates mass, charge, and intensity of detected ions. Correlating molecular weight with a given output peak can be accomplished directly from the spectral data, i.e., where the charge on an ion is one and the molecular weight is therefore equal to the numerator value minus the mass of the ionizing proton. However, protein ions can be complexed with various counter-ions and adducts, such as N, C, and K'. In such a case, it would be expected that a given protein ion would exhibit multiple peaks, such as a triplet, representing different ionic states (or species) of the

same protein. Thus, it may be necessary to analyze and process spectral data to determine families of peaks arising from the same protein. This analysis can be carried out conventionally, e.g., as described by Mann et al., anal. Chem., 61:1702-1708 (1989).

In matching a molecular mass calculated from a mass spectrometer to a molecular mass predicted from a database, such as a genomic or expressed gene database, post-translation processing may have to be considered. There are various processing events which modify protein structure, including, proteolytic processing, removal of N-terminal methionine, acetylation, methylation, glycosylation, phosphorylation, etc.

A database can be queried for a range of proteins matching the molecular mass of the unknown. The range window can be determined by the accuracy of the instrument, the method by which the sample was prepared, etc. Based on the number of hits (where a hit is match) in the spectrum, the unknown protein or peptide is identified or classified.

Methods of identifying one or more CPP by mass spectrometry are useful for diagnosis and prognosis of cardiovascular disorders. Preferably, such methods are used to detect one or more CPP present in human plasma. Exemplary techniques are described in U.S. Patent Applications 02/0060290, 02/0137106, 02/0138208, 02/0142343, 02/0155509, disclosures of which are incorporated by reference in their entireties.

Diagnostic and Prognostic Uses

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics; and in drug screening as further described herein.

The invention provides diagnostic and prognostic assays for detecting CPP nucleic acids and proteins, as further described. Also provided are diagnostic and prognostic assays for detecting interactions between CPPs and CPP target molecules, particularly natural agonists and antagonists.

The present invention provides methods for identifying polypeptides that are differentially expressed between two or more samples. "Differential expression" refers to differences in the quantity or quality of a polypeptide between samples. Such differences could result at any stage of protein expression from transcription through post-translational modification. For example, using protein array methods, two samples are bound to affinity spots on different sets of adsorbents (e.g., chips) and recognition maps are compared to identify polypeptides that are differentially retained by the two sets of adsorbents. Differential retention includes quantitative retention as well as qualitative differences in the polypeptide. For example, differences in post-translational modification of a protein can result

in differences in recognition maps detectable as differences in binding characteristics (e.g., glycosylated proteins bind differently to lectin adsorbents) or differences in mass (e.g., post-translational cleavage products). In certain embodiments, an adsorbent can have an array of affinity spots selected for a combination of markers diagnostic for a disease or syndrome.

Differences in polypeptide levels between samples (e.g., differentially expressed CPPs in plasma samples) can be identified by exposing the samples to a variety of conditions for analysis by desorption spectrometry (e.g., mass spectrometry). Unknown proteins can be identified by detecting physicochemical characteristics (e.g., molecular mass), and this information can be used to search databases for proteins having similar profiles.

Preferred methods of detecting a CPP utilize mass spectrometry techniques. Such methods provide information about the size and character of the particular CPP isoform that is present in a sample, e.g., a biological sample submitted for diagnosis or prognosis. Mass spectrometry techniques are detailed in the section titled "Detection of CPPs by mass spectrometry". Example 1 outlines a preferred detection scheme, wherein a biological sample is separated by chromatography before characterization by mass spectrometry. The invention provides a method of detecting a CPP in a biological sample comprising the steps of: fractionating a biological sample (e.g., plasma, serum, lymph, cerebrospinal fluid, cell lysate of a particular tissue) by at least one chromatographic step;

observed in mass spectrometry with known characteristics of CPP polypeptides (e.g., CPP 2, CPP 9, CPP 17, CPP 20 and CPP 21, as disclosed in Table 1).

The isolated nucleic acid molecules of the invention can be used, for example, to detect CPP mRNA (e.g., in a biological sample) or a genetic alteration in a CPP-encoding gene, and to modulate a CPP activity, as described further below. In addition, the CPPs can be used to screen for naturally occurring CPP target molecules, and to screen for drugs or compounds which modulate CPP activity. Moreover, the anti- CPP antibodies of the invention can be used to detect and isolate CPPs, regulate the bioavailability of CPPs, and modulate CPP activity.

Accordingly one embodiment of the present invention involves a method of use wherein a molecule of the present invention (e.g., a CPP, CPP nucleic acid, or CPP modulator) is used, for example, to diagnose, and/or prognose a disorder in which any of the aforementioned CPP activities is indicated. In another embodiment, the present invention involves a method of use wherein a molecule of the present invention is used, for example, for the diagnosis, and/or prognosis of subjects, preferably a human subject, in which any of the aforementioned activities is pathologically perturbed.

For example, the invention encompasses a method of determining whether a CPP is expressed

within a biological sample comprising: a) contacting said biological sample with: i) a polynucleotide that hybridizes under stringent conditions to a CPP nucleic acid; or ii) a detectable polypeptide (e.g. antibody) that selectively binds to a CPP; and b) detecting the presence or absence of hybridization between said polynucleotide and an RNA species within said sample, or the presence or absence of binding of said detectable polypeptide to a polypeptide within said sample. Detection of said hybridization or of said binding indicates that said CPP is expressed within said sample. Preferably, the polynucleotide is a primer, and said hybridization is detected by detecting the presence of an amplification product comprising said primer sequence, or the detectable polypeptide is an antibody.

In certain embodiments, detection involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202, the disclosures of which are incorporated herein by reference in their entireties), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegren et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1994) PNAS 91:360-364, the disclosures of which are incorporated herein by reference in their entireties), the latter of which can be particularly useful for detecting point mutations in the CPP-encoding-gene (see Abravaya et al. (1995) Nucleic Acids Res. 23:675-682, the disclosure of which is incorporated herein by reference in its entirety).

Also envisioned is a method of determining whether a mammal, preferably human, has an

said mammal; and b) comparing the amount of a CPP or of a CPP RNA species encoding a CPP within said biological sample with a level detected in or expected from a control sample. An increased amount of said CPP or said CPP RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has an elevated level of CPP expression, and a decreased amount of said CPP or said CPP RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has a reduced level of expression of a CPP.

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic purposes. Accordingly, one aspect of the present invention relates to diagnostic assays for determining CPP and/or nucleic acid expression as well as CPP activity, in the context of a biological sample (e.g., blood, plasma, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant CPP expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with a CPP, nucleic acid expression or

activity. For example, mutations in a CPP-encoding gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with CPP expression or activity.

The term "biological sample" is intended to include tissues, cells and biological fluids isolated from an individual, as well as tissues, cells and fluids present within an individual. That is, the detection methods of the invention can be used to detect a CPP mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. Preferred biological samples are biological fluids such as lymph, cerebrospinal fluid, blood, and especially blood plasma. For example, in vitro techniques for detection of a CPP mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of a CPP include mass spectrometry, Enzyme Linked Immuno Sorbent Assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection of a CPP-encoding genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of a CPP include introducing into an individual a labeled anti- CPP antibody.

For an example of immunoassays detecting the CPP 9 polypeptides of the invention, see Kurokawa, E. et al., Clin.Chim.Acta (1983), 128(1):83-93.

In preferred embodiments, the subject methods can be characterized by generally comprising detecting, in a tissue sample of the individual (e.g. a human patient), the presence or absence of a

CPP or (ii) the mis-expression of a CPP-encoding gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from the CPP-encoding gene, (ii) an addition of one or more nucleotides to the gene, (iii) a substitution of one or more nucleotides of the gene, (iv) a gross chromosomal rearrangement or amplification of the gene, (v) a gross alteration in the level of a messenger RNA transcript of the gene, (vi) aberrant modification of the gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene, and (viii) reduced level of expression, indicating lesion in regulatory element or reduced stability of a CPP-encoding transcript.

In yet another exemplary embodiment, aberrant methylation patterns of a CPP nucleic acid can be detected by digesting genomic DNA from a patient sample with one or more restriction endonucleases that are sensitive to methylation and for which recognition sites exist in the CPP-encoding gene (including in the flanking and intronic sequences). See, for example, Buiting et al. (1994) Human Mol Genet 3:893-895. Digested DNA is separated by gel electrophoresis, and hybridized with probes derived from, for example, genomic or cDNA sequences. The methylation status of the CPP-encoding gene can be determined by comparison of the restriction pattern generated

from the sample DNA with that for a standard of known methylation.

In yet another embodiment, a diagnostic assay is provided which detects the ability of a CPP to bind to a cell surface or extracellular protein. For instance, it will be desirable to detect CPP mutants which, while expressed at appreciable levels in the cell, are defective at binding a CPP target protein (having either diminished or enhanced binding affinity for the target). Such mutants may arise, for example, from mutations, e.g., point mutants, which may be impractical to detect by the diagnostic DNA sequencing techniques or by the immunoassays described above. The present invention accordingly further contemplates diagnostic screening assays which generally comprise cloning one or more CPP-encoding gene from the sample tissue, and expressing the cloned genes under conditions which permit detection of an interaction between that recombinant gene product and a target protein. As will be apparent from the description of the various drug screening assays set forth herein, a wide variety of techniques can be used to determine the ability of a CPP to bind to other components. These techniques can be used to detect mutations in a CPP-encoding gene which give rise to mutant proteins with a higher or lower binding affinity for a CPP target protein relative to the wild-type CPP. Conversely, by switching which of the CPP target protein and CPP is the "bait" and which is derived from the patient sample, the subject assay can also be used to detect CPP target protein mutants which have a higher or lower binding affinity for a CPP relative to a wild type form of that CPP target

In an exemplary embodiment, a target protein can be provided as an immobilized protein (a "target"), such as by use of GST fusion proteins and glutathione treated microtiter plates as described herein.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a CPP, mRNA, or genomic DNA, such that the level of a CPP, mRNA or genomic DNA is measured in the biological sample, and comparing the level of a CPP, mRNA or genomic DNA in the control sample with the level of a CPP, mRNA or genomic DNA in the test sample. The invention also encompasses kits for detecting the presence of a CPP, mRNA or genomic DNA in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting a CPP, mRNA or genomic DNA in a biological sample; means for determining the amount of a CPP in the sample; and means for comparing the amount of CPP in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect CPP or nucleic acid.

CPPs clusters

In one aspect of the invention, methods for the diagnosis of cardiovascular disorders comprise detecting in a test biological sample the presence or level of one or more CPP of the invention in combination with the detection of other Cardiovascular disorder Plasma Polypeptides (CPPs). Particularly preferred other CPPs for use in the diagnosis of cardiovascular disorders in combination with CPPs of the invention are listed in Table 2.

Table 2

Table 2

CPP #	CEX	RP1	Tryptic Sequences (RP2)
CPP 8	18	6	CLDPVDTPNPTR (7-8), YKKPECQSDWQCPGK (8)
CPP 8	18	7	CLDPVDTPNPTR (7-8)
CPP 8	18	11	CLDPVDTPNPTR (5)
CPP 8	18	13	CLDPVDTPNPTR (3)
CPP 12	10	8	QSGEDNQDLAISFAGNGLSALR (8-9)
CPP 12	11	8	ESLSGVCEISGR (9), QSGEDNQDLAISFAGNGLSALR (9)
CPP 12	11	9	QSGEDNQDLAISFAGNGLSALR (9)
CPP 12	11	10	QSGEDNQDLAISFAGNGLSALR (7)
CPP 12	11	11	ESLSGVCEISGR (7)
CPP 13	13	14	VSAQQVQGVHAR (9, 12)
CPP 13	13	18	FPVYDYDPSSLR (5), VNSQSLSPYLFR (5-6)
CPP 13	13	20	DYYVSTAVCR (4-5), FPVYDYDPSSLR (5), VNSQSLSPYLFR (4-5), VSAQQVQGVHAR (5)
CPP 13	13	21	DYYVSTAVCR (5), VNSQSLSPYLFR (5)
CPP 13	13	22	DYYVSTAVCR (10), FPVYDYDPSSLR (3), VNSQSLSPYLFR (3-4)
CPP 13	13	23	VNSQSLSPYLFR (3)
CPP 13	13	25	DYYVSTAVCR (1), VNSQSLSPYLFR (1)
CPP 13	14	13	DALSASVVK (15), DSGEDPATCAFQR (15), FPVYDYDPSSLR (15), VNSQSLSPYLFR (14), VSAQQVQGVHAR (15)
CPP 13	14	15	DSGEDPATCAFQR (10), VNSQSLSPYLFR (10)
CPP 13	14	19	VSAQQVQGVHAR (7)
CPP 13	14	21	VSAQQVQGVHAR (5, 7, 8)
CPP 13	14	22	VNSQSLSPYLFR (3)
CPP 13	14	25	VSAQQVQGVHAR (2)
CPP 13	15	13	VNSQSLSPYLFR (17-18)
CPP 13	15	15	VNSQSLSPYLFR (11)
CPP 13	18	22	VSAQQVQGVHAR (3)
CPP 14	15	4	GVSLRPIGASCR (10)
CPP 14	16	6	GVSLRPIGASCR (9)
CPP 14	17	5	GVSLRPIGASCRDDSECITR (9-10)
CPP 14	18	7	GVSLRPIGASCR (9)
CPP 15	2	7	LQCYNCPNPTADCK (24)
CPP 15	2	8	AGLQVYNK (15), LQCYNCPNPTADCK (15, 17, 18), LRENELTYYCCK (16-18)

Table 2

CPP #	CEX	RP1	Tryptic Sequences (RP2)
CPP 15	2	9	ENELTYYCCK (17), FEHCNFNDVTTR (17), LQCYNCPNPTADCK (17), LRENELTYYCCK (16-17)
CPP 15	2	10	LQCYNCPNPTADCK (12), LRENELTYYCCK (12)
CPP 15	3	9	FEHCNFNDVTTR (15-16), LQCYNCPNPTADCK (15-16)
CPP 15	3	10	AGLQVYNK (9, 11), FEHCNFNDVTTR (10-11), LQCYNCPNPTADCK (8-11, 16), LRENELTYYCCK (9-11, 13)
CPP 15	3	11	FEHCNFNDVTTR (9-10), LQCYNCPNPTADCK (9-11), LRENELTYYCCK (9-11)
CPP 15	3	12	LQCYNCPNPTADCK (7)
CPP 15	3	13	LQCYNCPNPTADCK (7-8)
CPP 15	4	9	FEHCNFNDVTTR (14-15), LQCYNCPNPTADCK (14-15), LRENELTYYCCK (14-15)
CPP 15	4	10	FEHCNFNDVTTR (9-11), LQCYNCPNPTADCK (9-10), LRENELTYYCCK (10)
CPP 15	5	11	AGLQVYNK (10), FEHCNFNDVTTR (10)
CPP 15	6	9	LQCYNCPNPTADCK (14)
CPP 15	6	10	FEHCNFNDVTTR (8)
CPP 15	6	11	TAVNCSSDFDACLITK (10)
CPP 15	7	10	FEHCNFNDVTTR (9)
CPP 16	10	15	CLTTDEYDGHSTYPSHQYQ (12), TVAGQDAVIVLLGTR (10), YVAVMPPHIGDQPLTGAYTVTLDGR (11)
CPP 16	10	16	CLTTDEYDGHSTYPSHQYQ (9), LQAVTDDHIR (9), YVAVMPPHIGDQPLTGAYTVTLDGR (9)
CPP 16	10	19	NDLSPTTVMSEGAR (7)
CPP 16	11	16	TVAGQDAVIVLLGTR (9)
CPP 16	13	17	NDLSPTTVMSEGAR (10)
CPP 18	10	11	QCIHQLCFTSLR (15-19)
CPP 18	11	6	LPPCENVDLQRPNGL (13)
CPP 18	11	7	SNYFRLPPCENVDLQRPNGL (13)
CPP 18	11	10	QCIHQLCFTSLR (14-15)
CPP 18	11	11	QCIHQLCFTSLR (12-17, 19, 20)
CPP 18	12	7	LPPCENVDLQRPNGL (12), SNYFRLPPCENVDLQRPNGL (9-11)
CPP 18	12	10	QCIHQLCFTSLR (13, 14)
CPP 18	12	11	LYSVHRPVK (11), QCIHQLCFTSLR (11)
CPP 18	13	7	LPPCENVDLQRPNGL (15), SNYFRLPPCENVDLQRPNGL (15)
CPP 19	3	7	MSSSYPTGLADVK (10)
CPP 19	4	7	AGPAQTLIRPQDMK (10), MSSSYPTGLADVK (10), MSSSYPTGLADVKAGPAQTLIRPQDMK (10)
CPP 40	17	20	EDPTVSALLTSEK (9), VPSLVGSFIR (8-9)
CPP 40	17	22	VPSLVGSFIR (7)
CPP 40	18	20	VPSLVGSFIR (9)
CPP 41	12	12	LQNNENNISCVER (9), STDTSCVNPPTVQNAHILSR (9)
CPP 41	14	12	TGESAEFVCK (9)
CPP 41	15	11	ITCTEEGWSPTPK (15-16), STDTSCVNPPTVQNAHILSR (15-16), TGESAEFVCKR (16)
CPP 41	16	29	YKPFSQVPTGEVFYYSCEYNFVSPSK (1)

Table 2

CPP #	CEX	RP1	Tryptic Sequences (RP2)
CPP 41	17	12	CLHPCVISR (9), EIMENYNIALR (9), INHGILYDEEK (9), ITCTEEGWSPTPK (9), LQNNENNISCVER (9), SFWTRITCTEEGWSPTPK (9), STDTSCVNPPTVQNAHILSR (9), TGESAEFVCKR (9), TTCWDGKLEYPTCAK (9)
CPP 41	17	13	CLHPCVISR (9), EATFCDFPK (9), EIMENYNIALR (9), GWSTPPK (9), INHGILYDEEK (9), ITCTEEGWSPTPK (9), LQNNENNISCVER (9), STDTSCVNPPTVQNAHILSR (9), TTCWDGKLEYPTCAK (9)
CPP 41	17	14	EIMENYNIALR (7), INHGILYDEEK (7), ITCTEEGWSPTPK (7), LQNNENNISCVER (7), STDTSCVNPPTVQNAHILSR (7), YKPFSQVPTGEVFYYSCEYNFVSPSK (7)
CPP 41	17	16	EATFCDFPK (5), EIMENYNIALR (5), LQNNENNISCVER (5), STDTSCVNPPTVQNAHILSR (5), TGESAEFVCK (5), TTCWDGKLEYPTCAK (5)
CPP 41	17	18	CLHPCVISR (4), EATFCDFPK (4), EIMENYNIALR (4), INHGILYDEEK (4), ITCTEEGWSPTPK (4), LEYPTCAK (4), SFWTR (4), STDTSCVNPPTVQNAHILSR (4), TTCWDGKLEYPTCAK (4)
CPP 41	17	20	EATFCDFPK (3), EIMENYNIALR (3), INHGILYDEEK (3), ITCTEEGWSPTPK (3), LQNNENNISCVER (3), STDTSCVNPPTVQNAHILSR (3), TGESAEFVCKR (3), TTCWDGKLEYPTCAK (3), YKPFSQVPTGEVFYYSCEYNFVSPSK (3)
CPP 41	17	22	EATFCDFPK (1), EIMENYNIALR (1), INHGILYDEEK (1), ITCTEEGWSPTPK (1), NGQWSEPPKCLHPCVISR (1), SFWTRITCTEEGWSPTPK (1), STDTSCVNPPTVQNAHILSR (1), TTCWDGKLEYPTCAK (1)
CPP 41	17	26	CLHPCVISR (1), EATFCDFPK (1), INHGILYDEEK (1), ITCTEEGWSPTPK (1), NGQWSEPPK (1), STDTSCVNPPTVQNAHILSR (1), TGESAEFVCK (1), [illegible] ITCTEEGWSPTPK (1), STDTSCVNPPTVQNAHILSR (1), TGESAEFVCKR (1)
CPP 41	17	27	CLHPCVISR (1), EATFCDFPK (1), EIMENYNIALR (1), INHGILYDEEK (1), ITCTEEGWSPTPK (1), LEYPTCAK (1), LQNNENNISCVER (1), SFWTR (1), STDTSCVNPPTVQNAHILSR (1)
CPP 41	17	29	CLHPCVISR (1), EATFCDFPK (1), EIMENYNIALR (1), INHGILYDEEK (1), ITCTEEGWSPTPK (1), LQNNENNISCVER (1), STDTSCVNPPTVQNAHILSR (1), TGESAEFVCK (1), TTCWDGKLEYPTCAK (1), YKPFSQVPTGEVFYYSCEYNFVSPSK (1)
CPP 41	17	30	CLHPCVISR (1), EIMENYNIALR (1), INHGILYDEEK (1), ITCTEEGWSPTPK (1), LQNNENNISCVER (1), STDTSCVNPPTVQNAHILSR (1), TTCWDGKLEYPTCAK (1)
CPP 41	18	12	CLHPCVISR (9-10), EATFCDFPK (9-10), EIMENYNIALR (9), INHGILYDEEK (9-10), ITCTEEGWSPTPK (9), LQNNENNISCVER (9-10), STDTSCVNPPTVQNAHILSR (9-10), TGESAEFVCKR (9), TTCWDGKLEYPTCAK (9-10)
CPP 41	18	18	EATFCDFPK (4), EIMENYNIALR (4), LQNNENNISCVER (4), STDTSCVNPPTVQNAHILSR (4), TTCWDGKLEYPTCAK (4)
CPP 41	18	19	EATFCDFPK (4), EIMENYNIALR (4), INHGILYDEEK (4), ITCTEEGWSPTPK (4), LQNNENNISCVER (4), NGQWSEPPK (4), STDTSCVNPPTVQNAHILSR (4)
CPP 41	18	20	LQNNENNISCVER (3), STDTSCVNPPTVQNAHILSR (3), TGESAEFVCKR (3), TTCWDGKLEYPTCAK (3)
CPP 41	18	22	EATFCDFPK (1), INHGILYDEEK (1), ITCTEEGWSPTPK (1), LQNNENNISCVER (1), STDTSCVNPPTVQNAHILSR (1), TGESAEFVCKR (1), TTCWDGKLEYPTCAK (1)

Table 2

CPP #	CEX	RP1	Tryptic Sequences (RP2)
CPP 41	18	26	CLHPCVISR (1), EATFCDFPK (1), EIMENYNIALR (1), INHGILYDEEK (1), ITCTEEGWSPTPK (1), STDTSCVNPPTVQNAHILSR (1), TGESAEFVCK (1)
CPP 41	18	29	EATFCDFPK (1), EIMENYNIALR (1), INHGILYDEEK (1), ITCTEEGWSPTPK (1), LQNNENNISCVER (1), STDTSCVNPPTVQNAHILSR (1), TTCWDGKLEYPTCAK (1)
CPP 149	13	9	AFTECCVVASQLR (11), CCYDGACVNNDETCEQR (10)
CPP 149	14	9	AFTECCVVASQLR (11), CCYDGACVNNDETCEQR (11)
CPP 149	15	9	AFTECCVVASQLR (8), CCYDGACVNNDETCEQR (8)
CPP 149	15	11	AFTECCVVASQLR (7)
CPP 149	16	8	AFTECCVVASQLR (8-9)
CPP 149	16	9	AFTECCVVASQLR (9)
CPP 149	17	10	AFTECCVVASQLR (7), CCYDGACVNNDETCEQR (7)
CPP 150	14	22	TNFDNDIALVR (12)
CPP 150	14	23	SNALDIIFQTDLTGQK (11), SSNNPHSPIVEEFQVPYNK (11), TNFDNDIALVR (10-11), VEDPESTLFGSVIR (8)
CPP 150	14	24	DVVQITCLDGFEVVEGR (12), EDTPNSVWEPAK (9), QFGPYCGHGFPGPLNIETK (8-12), SNALDIIFQTDLTGQK (8-9, 12), SSNNPHSPIVEEFQVPYNK (9), TNFDNDIALVR (8-12)
CPP 150	14	25	QFGPYCGHGFPGPLNIETK (9), SNALDIIFQTDLTGQK (5), TNFDNDIALVR (5, 9), VEDPESTLFGSVIR (9)
CPP 150	14	26	GDSGGAFAVQDPNDK (8), SNALDIIFQTDLTGQK (7-8), TNFDNDIALVR (7)
CPP 150	14	27	GDSGGAFAVQDPNDK (8), QFGPYCGHGFPGPLNIETK (5), SNALDIIFQTDLTGQK (6, 8), TNFDNDIALVR (5, 7, 8)
CPP 150	14	29	QFGPYCGHGFPGPLNIETK (7, 11), SSNNPHSPIVEEFQVPYNK (8), TNFDNDIALVR (7-8, 10, 11, 12, 13, 14, 16, 17, 19, 20)
CPP 150	14	30	GDSGGAFAVQDPNDK (8), QFGPYCGHGFPGPLNIETK (11), TNFDNDIALVR (7)
CPP 151	10	22	SFEGLGQLEVLTLDHNQLQEVK (8)
CPP 151	12	21	SFEGLGQLEVLTLDHNQLQEVK (14)
CPP 151	13	23	LAELPADALGPLQR (12), LAYLQPALFSGLAELR (11), LEALPNSLLAPLGR (12), VAGLLEDTFPGLLGLR (11-12)
CPP 501	2	21	EFLEDTCVQYVQK (7), TQSGLQSYLLQFHGLVR (7)
CPP 501	2	22	CFLGCELPPEGSR (6), EFLEDTCVQYVQK (6), TQSGLQSYLLQFHGLVR (4, 6)
CPP 501	2	23	EFLEDTCVQYVQK (6), TQSGLQSYLLQFHGLVR (5)
CPP 502	15	2	ALNSIIDVYHK (1)
CPP 502	17	17	ALNSIIDVYHK (9), GADVWFK (9)
CPP 502	17	18	ALNSIIDVYHK (7), LLETECPQYIR (7)
CPP 502	18	16	ALNSIIDVYHK (8)
CPP 502	18	17	ALNSIIDVYHK (8-10), GADVWFK (9), GNFHAVYR (9), LLETECPQYIR (8, 10), MLTELEK (9)
CPP 502	18	18	ALNSIIDVYHK (6-7), GADVWFK (7), LLETECPQYIR (6-7)
CPP 502	18	19	ALNSIIDVYHK (7), LLETECPQYIR (7)
CPP 502	18	20	ALNSIIDVYHK (6), LLETECPQYIR (5)
CPP 502	18	21	LLETECPQYIR (6)
CPP 502	18	22	ALNSIIDVYHK (5), LLETECPQYIR (4)

Table 2

CPP #	CEX	RP1	Tryptic Sequences (RP2)
CPP 502	18	23	LLETECPQYIR (5)
CPP 502	18	24	ALNSIIDVYHK (3), LLETECPQYIR (3)
CPP 502	18	25	ALNSIIDVYHK (3), LLETECPQYIR (3)
CPP 502	18	26	LLETECPQYIR (4)
CPP 502	18	27	ALNSIIDVYHK (4), LLETECPQYIR (4)
CPP 503	16	21	FALLGDFFR (6)
CPP 503	17	18	CMGTVTLNQAR (7), FALLGDFFR (6-8), GSFDISCDK (7)
CPP 503	17	19	CMGTVTLNQAR (7), FALLGDFFR (6-7), IKDFLR (7)
CPP 503	17	20	FALLGDFFR (6), FALLGDFFRK (6)
CPP 503	17	21	FALLGDFFR (5-6), IKDFLR (6), TTQQSPEDCDFK (6)
CPP 503	17	22	CMGTVTLNQAR (4), FALLGDFFR (4, 6)
CPP 503	17	23	CMGTVTLNQAR (4), FALLGDFFR (4), QVLSYKEAVLR (4)
CPP 503	17	24	CMGTVTLNQAR (2-3), FALLGDFFR (3)
CPP 503	17	25	CMGTVTLNQAR (2), FALLGDFFR (3), FALLGDFFRK (2)
CPP 503	17	26	TTQQSPEDCDFKK (2)
CPP 503	17	27	CMGTVTLNQAR (4), FALLGDFFR (4), TTQQSPEDCDFKK (4)
CPP 503	17	28	FALLGDFFR (5)
CPP 503	17	29	CMGTVTLNQAR (4), FALLGDFFR (4), GSFDISCDK (4), TTQQSPEDCDFK (4), TTQQSPEDCDFKK (4)
CPP 503	17	30	FALLGDFFR (4)
CPP 503	18	12	AIDGINQR (10)
CPP 503	18	19	CMGTVTLNQAR (7), FALLGDFFR (6-7), FALLGDFFRK (7)
CPP 503	18	20	CMGTVTLNQAR (5-6), FALLGDFFR (5-6), FALLGDFFRK (5-6)
CPP 503	18	21	CMGTVTLNQAR (6), FALLGDFFR (5-7), IKDFLR (5)
CPP 503	18	22	CMGTVTLNQAR (4), FALLGDFFR (4-5), GSFDISCDK (4), GSFDISCDKDNK (4), IKDFLR (4)
CPP 503	18	23	FALLGDFFR (4-5)
CPP 503	18	24	CMGTVTLNQAR (2), FALLGDFFR (2), FALLGDFFRK (2)
CPP 503	18	25	CMGTVTLNQAR (2), FALLGDFFR (2-3), FALLGDFFRK (2-3), GSFDISCDK (3)
CPP 503	18	26	CMGTVTLNQAR (4), FALLGDFFR (3-4), QVLSYKEAVLR (4)
CPP 503	18	27	CMGTVTLNQAR (4), FALLGDFFR (3-4), TTQQSPEDCDFK (4)
CPP 503	18	28	FALLGDFFR (3)
CPP 503	18	29	FALLGDFFR (3-4)
CPP 503	18	30	CMGTVTLNQAR (3-4), FALLGDFFR (3-4), FALLGDFFRK (3), TTQQSPEDCDFKK (4)
CPP 504	9	15	VPLQQNFQDNQFQGK (15)
CPP 504	9	16	CDYWIR (11), ELTSELK (10), MYATIYELK (10-11), SLGLPENHIVFPVPIDQCIDG (10), SYPGLTSYLVR (11), TFVPGCQPGEFTLGNIK (10-11), VPLQQNFQDNQFQGK (10-11), VVSTNYNQHAMVFFK (10), WYVVGLAGNAILR (10)
CPP 504	9	18	VPLQQNFQDNQFQGK (8)
CPP 505	6	8	VVEPPEKDDQLVVLFPVQKPK (8)
CPP 505	6	10	AWMETEDTLGR (8)
CPP 505	7	9	VVEPPEKDDQLVVLFPVQKPK (12)

Table 2

CPP #	CEX	RP1	Tryptic Sequences (RP2)
CPP 505	8	10	AWMETEDTLGR (8), VVEPPEKDDQLVVLFPVQKPK (8)
CPP 505	8	11	AWMETEDTLGR (8-9), HWPSEQDPEKAWGAR (8), LLTTEEKPR (8), LWVMPNHQVLLGPEEDQDHIYHPQ (8), VVEPPEKDDQLVVLFPVQKPK (8)
CPP 505	9	8	LLTTEEKPR (11-13), VVEPPEKDDQLVVLFPVQKPK (12)
CPP 505	9	9	AWMETEDTLGR (11, 13, 14, 15), GPILPGTK (13), HWPSEQDPEK (14), HWPSEQDPEKAWGAR (12), LLTTEEKPR (10, 13, 15, 17), VVEPPEKDDQLVVLFPVQKPK (11-15)
CPP 505	9	10	AWMETEDTLGR (9-11, 13, 14), DDQLVVLFPVQKPK (9-10), LWVMPNHQVLLGPEEDQDHIYHPQ (10), VVEPPEKDDQLVVLFPVQKPK (11)
CPP 505	9	11	AWMETEDTLGR (9-11, 13), GPILPGTK (11-12), HWPSEQDPEKAWGAR (11), LLTTEEKPR (10-12), LLTTEEKPRGQGR (11), LWVMPNHQVLLGPEEDQDHIYHPQ (11-12), VLSPEPDHDSLYHPPPEEDQGEERPR (11), VVEPPEKDDQLVVLFPVQKPK (10-11)
CPP 505	10	9	AWMETEDTLGR (15-18), LLTTEEKPR (14-16), VVEPPEKDDQLVVLFPVQKPK (17)
CPP 505	10	10	AWMETEDTLGR (8, 10, 11), HWPSEQDPEK (10), LLTTEEKPR (10), LWVMPNHQVLLGPEEDQDHIYHPQ (10-11), VVEPPEKDDQLVVLFPVQKPK (9-11)
CPP 505	10	11	AWMETEDTLGR (8), VVEPPEKDDQLVVLFPVQKPK (11)
CPP 505	11	9	LLTTEEKPR (14), VVEPPEKDDQLVVLFPVQKPK (16-17)
CPP 505	11	10	AWMETEDTLGR (11), LWVMPNHQVLLGPEEDQDHIYHPQ (11), VVEPPEKDDQLVVLFPVQKPK (9, 11)
CPP 506	8	19	EVMPSIQSLDALVK (5)
CPP 506	8	20	EVMPSIQSLDALVK (5)
CPP 506	8	24	EVMPSIQSLDALVK (5)
CPP 506	8	25	EVMPSIQSLDALVK (3)
CPP 507	5	7	NANTFISPQQR (11-12)
CPP 507	6	6	NANTFISPQQR (8)
CPP 507	8	8	NANTFISPQQR (7)
CPP 507	10	7	NANTFISPQQR (7-8, 11), YESHESMESYELNPFINRR (12)
CPP 507	10	8	NANTFISPQQR (6)
CPP 507	11	8	NANTFISPQQR (8-9, 11, 12)
CPP 507	11	11	NANTFISPQQR (8)
CPP 507	12	8	NANTFISPQQR (6-7, 12)
CPP 507	12	9	NANTFISPQQR (7-9)
CPP 507	13	8	NANTFISPQQR (8)
CPP 507	13	9	NANTFISPQQR (8-12)
CPP 507	14	6	NANTFISPQQR (10)
CPP 507	14	7	NANTFISPQQR (10-14)
CPP 507	14	9	NANTFISPQQR (8)
CPP 507	14	11	NANTFISPQQR (7)
CPP 507	15	7	NANTFISPQQR (16)
CPP 507	16	8	NANTFISPQQR (7)
CPP 508	4	12	GPETLCGAELVDALQFVCGDR (8)
CPP 508	5	13	GPETLCGAELVDALQFVCGDR (8-9)

Table 2

CPP #	CEX	RP1	Tryptic Sequences (RP2)
CPP 508	5	16	GPETLCGAELVDALQFVCGDR (4)
CPP 508	6	11	GPETLCGAELVDALQFVCGDR (10)
CPP 508	6	12	GPETLCGAELVDALQFVCGDR (8)
CPP 508	6	13	GPETLCGAELVDALQFVCGDR (8, 12)
CPP 508	7	10	GPETLCGAELVDALQFVCGDR (11)
CPP 508	7	11	GFYFNKPTGYGSSSR (11), GPETLCGAELVDALQFVCGDR (11-13), RAPQTGIVDECCFR (13-14)
CPP 508	7	12	GFYFNKPTGYGSSSR (9), GPETLCGAELVDALQFVCGDR (8-9)
CPP 508	8	12	APQTGIVDECCFR (8), GPETLCGAELVDALQFVCGDR (7-8)
CPP 508	8	13	GFYFNKPTGYGSSSR (6-7), GPETLCGAELVDALQFVCGDR (6-8), RAPQTGIVDECCFR (7)
CPP 508	8	14	APQTGIVDECCFR (6), GFYFNKPTGYGSSSR (6), GPETLCGAELVDALQFVCGDR (6)
CPP 508	9	11	APQTGIVDECCFR (11), GPETLCGAELVDALQFVCGDR (10-11)
CPP 508	9	12	APQTGIVDECCFR (1, 7, 8, 9), GFYFNKPTGYGSSSR (1, 7, 8), GPETLCGAELVDALQFVCGDR (1, 7, 8, 9, 10, 11, 12), RAPQTGIVDECCFR (7-8)
CPP 508	9	13	APQTGIVDECCFR (8-9), GPETLCGAELVDALQFVCGDR (8-9), RAPQTGIVDECCFR (8)
CPP 508	9	14	GFYFNKPTGYGSSSR (7), GPETLCGAELVDALQFVCGDR (7), RAPQTGIVDECCFR (6-7)
CPP 508	9	15	GFYFNKPTGYGSSSR (7), RAPQTGIVDECCFR (7)
CPP 508	9	16	GFYFNKPTGYGSSSR (5)
CPP 508	9	19	RAPQTGIVDECCFR (3)
CPP 508	10	12	APQTGIVDECCFR (10-12), GFYFNKPTGYGSSSR (8-10, 12), APQTGIVDECCFR (8-9), GFYFNKPTGYGSSSR (7-9), GPETLCGAELVDALQFVCGDR (7-9, 12), RAPQTGIVDECCFR (8), RLEMYCAPLKPAK (7)
CPP 508	10	13	GFYFNKPTGYGSSSR (7), GPETLCGAELVDALQFVCGDR (7-8, 12)
CPP 508	11	10	GFYFNKPTGYGSSSR (11), GPETLCGAELVDALQFVCGDR (9-11)
CPP 508	11	11	APQTGIVDECCFR (10-12), GFYFNKPTGYGSSSR (9-11), GPETLCGAELVDALQFVCGDR (9-14), LEMYCAPLKPAK (11), RAPQTGIVDECCFR (9-12), RLEMYCAPLKPAK (10)
CPP 508	11	12	APQTGIVDECCFR (7-8), GPETLCGAELVDALQFVCGDR (7-8)
CPP 508	11	13	APQTGIVDECCFR (8), GPETLCGAELVDALQFVCGDR (8-9)
CPP 508	11	19	GPETLCGAELVDALQFVCGDR (3)
CPP 508	12	11	APQTGIVDECCFR (8-10), GFYFNKPTGYGSSSR (8-9), GPETLCGAELVDALQFVCGDR (8-10), LEMYCAPLKPAK (9), RAPQTGIVDECCFR (9), RLEMYCAPLKPAK (9)
CPP 508	12	12	APQTGIVDECCFR (7-8), GFYFNKPTGYGSSSR (7, 10), GPETLCGAELVDALQFVCGDR (7-13), RAPQTGIVDECCFR (7)
CPP 508	12	13	APQTGIVDECCFR (12), GFYFNKPTGYGSSSR (11-12), GPETLCGAELVDALQFVCGDR (8, 12)
CPP 508	12	14	RAPQTGIVDECCFR (6)
CPP 508	12	19	GPETLCGAELVDALQFVCGDR (3)
CPP 508	12	20	GPETLCGAELVDALQFVCGDR (1)
CPP 508	13	11	GFYFNKPTGYGSSSR (8), GPETLCGAELVDALQFVCGDR (9)
CPP 508	13	12	GPETLCGAELVDALQFVCGDR (8)
CPP 508	13	13	GPETLCGAELVDALQFVCGDR (8-9)

Table 2

CPP #	CEX	RP1	Tryptic Sequences (RP2)
CPP 508	14	10	APQTGIVDECCFR (9), GFYFNKPTGYGSSSR (9), GPETLCGAELVDALQFVCGDR (8-10), RAPQTGIVDECCFR (8)
CPP 508	14	11	GPETLCGAELVDALQFVCGDR (9, 11)
CPP 508	14	12	APQTGIVDECCFR (8), GPETLCGAELVDALQFVCGDR (7-8)
CPP 508	14	13	GPETLCGAELVDALQFVCGDR (8)
CPP 508	15	10	APQTGIVDECCFR (8), GPETLCGAELVDALQFVCGDR (8-10)
CPP 508	15	11	APQTGIVDECCFR (8-9), GPETLCGAELVDALQFVCGDR (8-9)
CPP 508	15	12	GFYFNKPTGYGSSSR (7), GPETLCGAELVDALQFVCGDR (7-8, 10)
CPP 508	15	13	GPETLCGAELVDALQFVCGDR (8)
CPP 508	16	11	GPETLCGAELVDALQFVCGDR (9)
CPP 508	16	12	APQTGIVDECCFR (6), GFYFNKPTGYGSSSR (6), GPETLCGAELVDALQFVCGDR (6-8)
CPP 508	16	13	GPETLCGAELVDALQFVCGDR (7)
CPP 508	17	11	APQTGIVDECCFR (8), GFYFNKPTGYGSSSR (8), GPETLCGAELVDALQFVCGDR (7-8)
CPP 508	17	12	APQTGIVDECCFR (6), GFYFNKPTGYGSSSR (6), GPETLCGAELVDALQFVCGDR (6-7)
CPP 508	17	13	APQTGIVDECCFR (6), GFYFNKPTGYGSSSR (6), GPETLCGAELVDALQFVCGDR (6-7)
CPP 508	17	15	GPETLCGAELVDALQFVCGDR (5)
CPP 508	17	20	GPETLCGAELVDALQFVCGDR (1)
CPP 508	18	11	GPETLCGAELVDALQFVCGDR (8)
CPP 508	18	12	GPETLCGAELVDALQFVCGDR (6, 8)
CPP 508	18	13	GPETLCGAELVDALQFVCGDR (6-7)
CPP 508	18	18	GPETLCGAELVDALQFVCGDR (2)
CPP 509	6	6	VPFNGQDPVK (10)
CPP 509	6	8	AQEPVKGPVSTKPGSCPIILIR (7-8), CAMLNPPNR (7-8), CLKDTDCPGIK (7), VPFNGQDPVK (7), VPFNGQDPVKGQVSVK (7)
CPP 509	8	8	AQEPVKGPVSTKPGSCPIILIR (7), CAMLNPPNR (7)
CPP 509	9	8	VPFNGQDPVK (7)
CPP 509	10	8	GPVSTKPGSCPIILIR (8)

Table 2 details, for each CPP, the sequences detected by mass spectrometry according to the procedures described in Example 1. In addition, Table 2 indicates in which fractions of the CEX, RP1, and RP2 chromatographies each sequence was found.

The CPPs listed in Table 2 were all identified as differentially expressed between individuals with cardiovascular disorders and control individuals using the procedure described in Example 1. In particular, each CPP listed in Table 2 was found to vary between the control and disease samples as detailed in Table 3 below.

Table 3

Table 3	
CPP #	**Direction of variation**
CPP 8	Identified in Disease only
CPP 12	Identified at a higher level in Disease
CPP 13	Identified at a higher level in Controls
CPP 14	Identified at a higher level in Disease
CPP 15	Identified at a higher level in Disease
CPP 16	Identified at a higher level in Disease
CPP 18	Identified at a higher level in Disease
CPP 19	Identified in Controls only
CPP 40	Identified at a higher level in Controls
CPP 41	Identified at a higher level in Controls
CPP 149	Identified in Disease only
CPP 150	Identified at a higher level in Disease
CPP 151	Identified at a higher level in Disease
CPP 501	Identified at a higher level in Disease
CPP 502	Identified at a higher level in Controls
CPP 503	Identified at a higher level in Controls
CPP 504	Identified at a higher level in Controls
CPP 505	Identified at a higher level in Disease
CPP 506	Identified at a higher level in Disease
CPP 507	Identified at a higher level in Disease
CPP 508	Identified in Disease only

One skilled in the art can use CPPs of the invention with a number of additional CPPs from Table 2, chosen using a suitable analysis of the levels of the CPPs from Table 2 measured in a number of diseased individuals and control individuals through the methods of Example 1. The strategies for discovering such combinations of CPPs need to regard each CPP as one variable and the disease as a joint, multi-variate effect caused by interaction of these variables.

Linear Discriminant Analysis (LDA) is one such analysis procedure, which can be used to detect significant association between a cluster of variables (*i.e.* CPPs) and cardiovascular diseases. In performing LDA, a set of weights is associated with each variable (*i.e.* CPP) so that the linear combination of weights and the measured values of the variables can identify the disease state by discriminating between subjects having a cardiovascular diease and subjects free from cardivascular diseases. Enhancements to the LDA allow stepwise inclusion (or removal) of variables to optimize the discriminant power of the model. The results of the LDA is therefore a cluster of CPPs which can be used without limitations for diagnosis, prognosis, therapy or drug development. Other enhanced versions of LDA, such as Flexible Discriminant Analysis permit the use of non-linear combinations of

variables to discriminate a disease state from a normal state. The results of the discriminant analysis can be verified by *post-hoc* tests and also by repeating the analysis using alternative techniques such as classification trees.

Drug screening assays

The invention provides a method (also referred to herein as a "screening assay") for identifying candidate modulators (e.g., small molecules and peptides, antibodies, peptidomimetics or other drugs) which bind to CPPs, have a modulatory effect on, for example, CPP expression or preferably CPP biological activity. In some embodiments small molecules can be generated using combinatorial chemistry or can be obtained from a natural products library. Assays may be cell based or non-cell based assays. Drug screening assays may be binding assays or more preferentially functional assays, as further described.

When the invention is used for drug development, e.g., to determine the ability of a CPP modulator or drug candidate to induce an anti-cardiovascular disorder response, the body fluid analyzed for the level of at least one CPP is preferably from a non-human mammal. The non-human mammal is preferably one in which the induction of an anti-cardiovascular disorder response by endogenous and/or exogenous agents is predictive of the induction of such a response in a human.

Agents that are found, using screening assays as further described herein, to modulate CPP activity by at least 5%, more preferably by at least 10%, still more preferably by at least 30%, still more preferably by at least 50%, still more preferably by at least 70%, even more preferably by at least 90 %, may be selected for further testing as a prophylactic and/or therapeutic anti-cardiovascular disease agent.

In another aspect, agents that are found, using screening assays as further described herein, to modulate CPP expression by at least 5%, more preferably by at least 10%, still more preferably by at least 30%, still more preferably by at least 50%, still more preferably by at least 70%, even more preferably by at least 90 %, may be selected for further testing as a prophylactic and/or therapeutic anti-cardiovascular disease agent.

Agents that are found to modulate CPP activity may be used, for example, to modulate treatment regimens for cardiovascular disorders, or to reduce the symptoms of a cardiovascular disorder alone or in combination with other appropriate agents or treatments.

Protein array methods are useful for screening and drug discovery. For example, one member of a receptor/ ligand pair is docked to an adsorbent, and its ability to bind the binding partner is